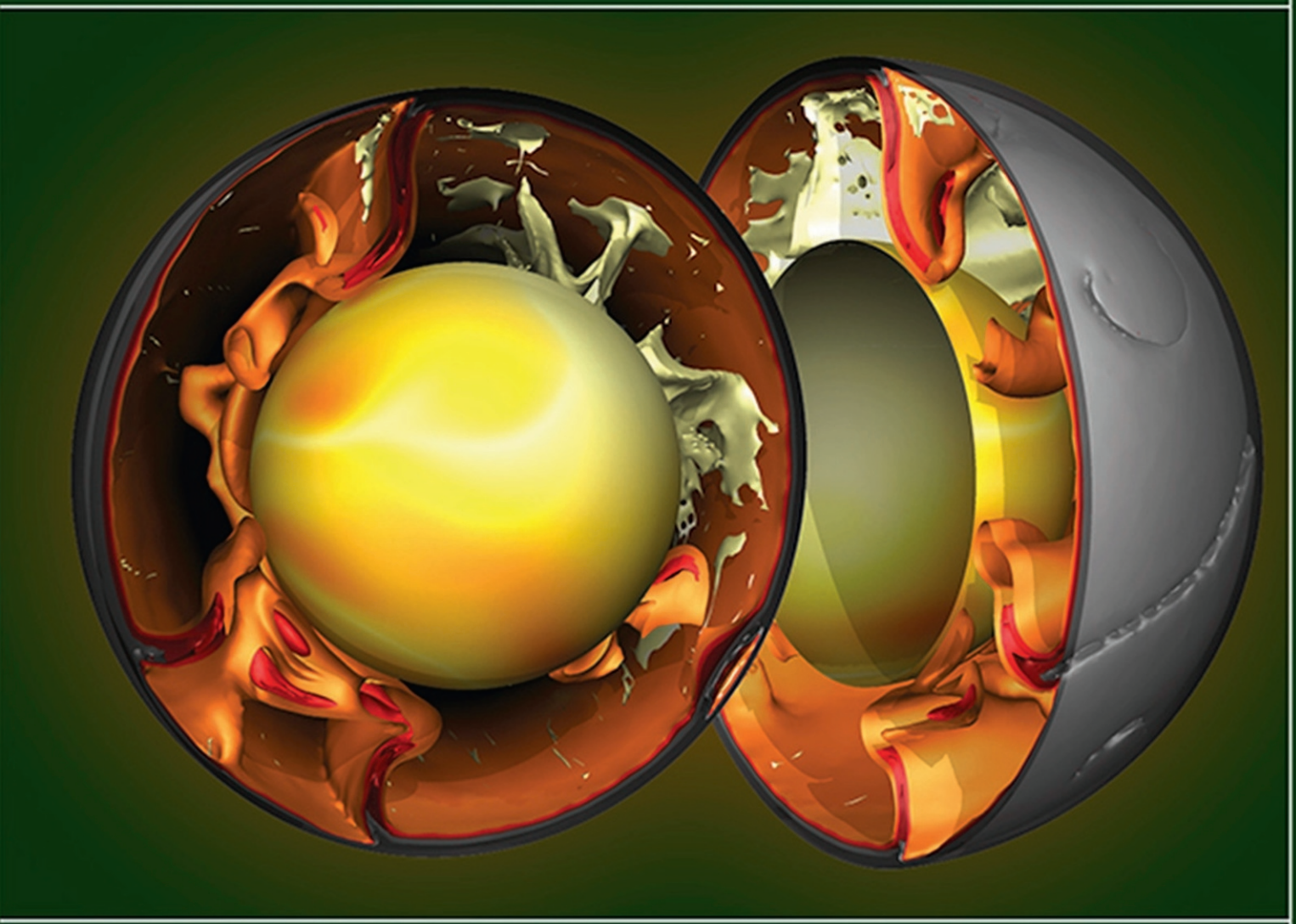
SECOND EDITION

TREATISE ON GEOPHYSICS

Volume 9

EVOLUTION OF THE EARTH



VOLUME EDITOR: DAVID STEVENSON

EDITOR-IN-CHIEF: GERALD SCHUBERT

TREATISE ON GEOPHYSICS

SECOND EDITION

This page intentionally left blank

TREATISE ON GEOPHYSICS

SECOND EDITION

EDITOR-IN-CHIEF

Gerald Schubert

Distinguished Professor Emeritus, Department of Earth, Planetary and Space Sciences, University of California, Los Angeles

VOLUME 9

EVOLUTION OF THE EARTH

VOLUME EDITOR

Professor David Stevenson

Division of Geological & Planetary Science, California Institute of Technology

ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington MA 01803, USA

Copyright © 2015 Elsevier B.V. unless otherwise stated. All rights reserved.

Cover image created for this work by Fabio Crameri, University College London.
The figure depicts the temperature field from a convection calculation using the code STAG-YY, written by Paul Tackley.

The following chapters are US Government works in the public domain and not subject to copyright:

1.11 Crust and Lithospheric Structure - Global Crustal Structure; 3.09 Earth Rotation Variations – Long Period; 4.07 The Role of Fault-Zone Drilling; 5.11 The Time-Averaged Field and Paleosecular Variation; 10.21 Exoplanetary Geophysics: An Emerging Discipline; 11.12 Tools and Techniques: Gravitational Method.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN (print): 978-0-444-53802-4

For information on all Elsevier publications
visit our website at http://store.elsevier.com/

Editorial & Production: Gemma Tomalin, Richard Berryman, Marise Willis, Michael Nicholls, Ellen Bishop
Cover design: Matthew Limbert

Printed and bound in the United Kingdom

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface *vii*

Editor-in-Chief *ix*

Contributors *xi*

Contents of all Volumes *xiii*

9.01 Evolution of the Earth: An Introduction and Overview 1
DJ Stevenson

9.02 The Composition and Major Reservoirs of the Earth Around the Time of the Moon-Forming Giant Impact 11
AN Halliday and BJ Wood

9.03 Formation of the Earth's Core 43
DC Rubie, F Nimmo, and HJ Melosh

9.04 Magma Oceans and Primordial Mantle Differentiation 81
V Solomatov

9.05 Water in the Evolution of the Earth and Other Terrestrial Planets 105
S Karato

9.06 Evolution of the Earth: Plate Tectonics Through Time 145
NH Sleep

9.07 Mechanism of Continental Crustal Growth 173
M Stein and Z Ben-Avraham

9.08 Thermal and Compositional Evolution of the Core 201
F Nimmo

9.09 The History of the Earth's Rotation: Impacts of Deep Earth Physics and Surface Climate Variability 221
WR Peltier

9.10 Coevolution of Life and Earth 281
GJ Retallack

This page intentionally left blank

PREFACE

The *Treatise on Geophysics* is a comprehensive and in-depth study of the physics of the Earth and of terrestrial-like planets residing both within and outside our solar system. Its breadth and detail of coverage are beyond what any single geophysics text can provide. The first edition of the *Treatise on Geophysics* was published in 2007, nearly a decade ago. Of course, there is much progress in science in that length of time and the field of geophysics has grown rapidly and developed new lines of inquiry. Accordingly, we have brought the *Treatise on Geophysics* up-to-date with this second edition. The new edition will continue to provide students and professionals with fundamental and state-of-the-art discussion of all aspects of geophysics. In this new edition, the reader will find updates to all the chapters contained in the first edition and new chapters that discuss topics missing from the first edition. In a few instances, chapters from the first edition have simply been reprinted in order to avoid gaps in coverage. Chapters that emphasized fundamental physics often required little updating.

A highlight of the second edition is a new volume on Resources in the Near-Surface Earth. The new volume discusses the role of geophysics in the exploitation and conservation of natural resources and the assessment of degradation of natural systems by pollution. The near surface is a zone where humans and natural systems interact. Understanding the effects of human impacts is a challenge particularly for the long term.

There are four new chapters in the volume on Deep Earth Seismology. One of them describes progress in seismic instrumentation in the last 30 years in the era of broadband digital sensor and recording systems. Another one addresses a new approach for the determination of crust and uppermost mantle structure, based on ambient noise cross-correlations. This has revolutionized the field, since it allows the study of regional structure at high resolution in areas where there are no earthquakes. The third new chapter is devoted to recent advances in characterizing the lithosphere–asthenosphere boundary under both oceans and continents. Finally, the last new chapter discusses the morphology of subducted slabs in the Earth's transition zone.

The Mineral Physics volume has new chapters on dynamic (shock-wave) compression methods, and on the mineralogy of the upper mantle and the D″ layer focusing on post-perovskite, and extrasolar super-Earths.

New material in the Geodesy volume expands the temporal resolution and latency of geodetic measurements to high-rate, real-time measurements and increases the spatial resolution to meter size pixels with optical and radar imaging systems. Application areas are extended with new material on the uses of Earth gravity measurements.

The Earthquake Seismology volume has a new chapter on rock mechanics that is a comprehensive account of modern rock mechanics with emphasis on the variation of frictional properties of rocks during an earthquake cycle. Many chapters have considerable updates reflecting the lessons from significant recent earthquakes (e.g., the 2011 Tohoku-Oki earthquake and the 2013 Okhotsk deep earthquake), new observations on slow slips and triggered earthquakes, and developments of new methods (e.g., back projection method). Some chapters are updated (e.g., paleoseismology and strong ground motion) with substantial new material.

Chapters in the Geomagnetism volume have been updated to include recent advances in the acquisition of larger data sets and higher quality data and in the construction of more reliable and sophisticated models. Improvements in data have occurred in areas such as magnetic anomalies, paleointensities, and planetary magnetism. Improvements in models of the present day field include so-called comprehensive models and models that incorporate data assimilation.

The Crustal and Lithosphere Dynamics volume has new contributions on mountain building and landscape evolution.

The Mantle Dynamics volume contains new chapters on the generation of plate tectonics from mantle convection and the core–mantle boundary region.

The Core Dynamics volume features a comprehensive update of this rapidly changing subject, including new chapters on turbulence and waves in the fluid outer core.

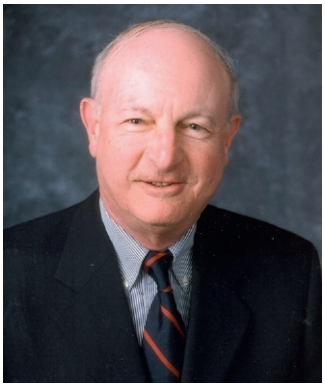
The volume on the Evolution of the Earth presents a new assessment on the role of volatiles in Earth evolution.

There are new chapters in the Physics of Terrestrial Planets and Moons volume on planetary atmospheres, planetary tectonics, extrasolar planets and asteroids and comets. The study of extrasolar planets has had explosive growth in the last decade and new spacecraft missions have significantly enhanced our knowledge of solar system planets and moons.

Finally, this is an opportunity for me to thank and acknowledge the editors of the individual volumes for their hard work, patience, dedication, and commitment in bringing this new edition of the Treatise to fruition. I am also indebted to Gemma Tomalin of Elsevier for her care and attention to the organization and production of the Treatise and to Deepa Arumugam for her careful coordination of the proofs from the copyeditor and typesetter. Of course, the authors of the individual chapters are the ones who are responsible for the excellence of the Treatise. We have all had some ups and downs in the process of producing the second edition of the Treatise but the final product is one that we can all view with pride.

Gerald Schubert
Editor-in-Chief

EDITOR-IN-CHIEF



Gerald Schubert is Distinguished Professor Emeritus in the Department of Earth, Planetary and Space Sciences at the University of California, Los Angeles. His research interests encompass the physics of the interiors and atmospheres of the Earth, the Moon, and the other moons and planets of the solar system. He is co-author with Donald Turcotte of *Geodynamics* (ed. 3, Cambridge University Press, 2014) and with Turcotte and Peter Olson of *Mantle Convection in the Earth and Planets* (Cambridge University Press, 2001). He is the author of over 540 research papers. He has participated in a number of NASA's planetary missions, including Apollo, Pioneer Venus, Magellan, and Galileo, and has been editor and editorial board member of many journals, including *Icarus*, *Journal of Geophysical Research*, *Geophysical Research Letters*, and *Annual Reviews of Earth and Planetary Sciences*. Professor Schubert is a Fellow of the American Geophysical Union and a recipient of the Union's James B. MacElwane medal and the Harry H. Hess medal. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

This page intentionally left blank

CONTRIBUTORS

Z Ben-Avraham
Tel Aviv University, Ramat Aviv, Israel

AN Halliday
University of Oxford, Oxford, UK

S Karato
Yale University, New Haven, CT, USA

HJ Melosh
Purdue University, West Lafayette, IN, USA

F Nimmo
University of California, Santa Cruz, CA, USA

WR Peltier
University of Toronto, Toronto, ON, Canada

GJ Retallack
University of Oregon, Eugene, OR, USA

DC Rubie
Universität Bayreuth, Bayreuth, Germany

NH Sleep
Stanford University, Stanford, CA, USA

V Solomatov
Washington University in St. Louis, St. Louis, MO, USA

M Stein
Geological Survey of Israel, Jerusalem, Israel

DJ Stevenson
California Institute of Technology, Pasadena, CA, USA

BJ Wood
University of Oxford, Oxford, UK

This page intentionally left blank

9.01 Evolution of the Earth: An Introduction and Overview

DJ Stevenson, California Institute of Technology, Pasadena, CA, USA

© 2015 Elsevier B.V. All rights reserved.

9.01.1	**Introduction**	1
9.01.1.1	How Should We Think of the Earth and Earth Evolution?	1
9.01.1.2	History and Themes	2
9.01.2	**Physical and Chemical Constraints**	4
9.01.2.1	Important Ideas	4
9.01.2.2	Some Useful Estimates	4
9.01.3	**Commentary on Formation Models**	6
9.01.4	**Commentary on Early Evolution Models**	7
9.01.5	**Outstanding Questions**	8
References		9

9.01.1 Introduction

9.01.1.1 How Should We Think of the Earth and Earth Evolution?

Evolutionary science is for the most part based on observation and indirect inference. It is not experimental science, even though experiments can certainly play a role in our understanding the processes. We can never hope to have the resources to build our own planet and observe how it evolves, but we can now hope to observe an ensemble of Earth-like planets elsewhere in the universe at diverse stages of their evolution (for a current state of exoplanet discovery, see exoplanets.org or exoplanet.eu). There are two central ideas that govern our thinking about the Earth and its history. One is *provenance*: The nature and origin of the material that went into making the Earth. This is our cosmic heritage, one that we presumably share with neighboring terrestrial planets and (to some uncertain extent) we share with the meteorites and the abundances of elements in the Sun. The other is *process*: The Earth is an engine and its current structure is a consequence of those ongoing processes, expressed in the form it takes now. The most obvious and important of these processes are plate tectonics and the inextricably entwined process of mantle convection. However, this central evolutionary process cannot be separated from the nature of the atmosphere and ocean, the geochemical evolution of various parts of the Earth expressed in the rock record, and life.

Figure 1 shows conceptually the ideas of Earth evolution, expressed as a curve in some multidimensional space that is here simplified by focusing on two variables ('this' and 'that'), the identities of which are not important. They could be physical variables such as temperature or chemical variables (composition of a particular reservoir) or isotopic tracers. The figure intends to convey the idea that we have an initial condition, an evolutionary path, and a present state. The initial condition is dictated not only by provenance but also by the physics of the formation process. By analogy, we would say that the apples from an apple tree owe much of their nature to not only the genetics of apples (the process of their formation) but also, to some extent, the soil and climate in which the tree grew. We are informed of this initial condition by astronomy, which tells us about how planets form in other solar systems, by geochemistry (a memory within the Earth of the materials and conditions of Earth formation) and by physical modeling: simulations and analysis of what may have occurred. Notably, we do not get information on the initial condition from geology since there are no rocks or landforms that date back to the earliest history of the Earth. Geology, aided by geochemistry and geobiology, plays a central role informing us about Earth history. Though some geophysicists study evolution, nearly all geophysical techniques are directed toward understanding a snapshot of the present Earth or a very short period prior to the present Earth, and it is only through modeling (e.g., of geologic data) that the physical aspects of evolution are illuminated.

In Figure 2, another important idea is conveyed: For many purposes, we should think of time logarithmically. This is in striking contrast to the way many geoscientists think of time, because they focus (naturally enough) on where the record is best. As a result, far more geologic investigation is carried out for the Phanerozoic (10% of Earth history) than the entire period before this. More importantly, the processes that govern early history are very energetic and fast. As a consequence, more could have happened in the first millions to hundreds of millions of years than throughout all of subsequent geologic time.

Table 1 develops this idea further by identifying some of the important timescales of relevance to Earth history and prehistory (here taken to mean the important events that took place even before the Earth formed). From this emerges the subdivision of geologic time into the accretion phase (the aggregation of bodies to make the Earth), lasting a hundred million years at most, an early evolution in which the high-energy consequences of the accretion (the stored heat) and possibly later impacts still play a role, perhaps lasting as long as half a billion years, and the rest of geologic time in which the energetics of the Earth is strongly affected by the long-lived radioactive heat sources.

In this overview chapter, an attempt is made to identify the main themes of Earth history, viewed geophysically, and to provide a context for appreciating the more detailed following chapters. At the end, some of the outstanding issues are revisited, reminding us that this is very much a living science in which there are many things not known or understood.

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00151-2

9.01.1.2 History and Themes

Hopkins (1839) in his *Preliminary observations on the refrigeration of the globe* illustrates well the prevailing view of that time that the Earth started hot and was cooling over time. This hot beginning now seems natural to us as a consequence of the gravitational energy of Earth formation, and it has been a consistently popular view even when the justifications for its advocacy were imperfectly developed. Famously, Lord Kelvin (Figure 3) took the hot initial Earth and applied conduction theory to the outermost region to estimate the age of the Earth at a hundred million years or less. Burchfield (1975) in his book, *Lord Kelvin and the age of the Earth*, documented Kelvin's various estimates and the conflict with the Victorian geologists of the time who believed that Kelvin's estimates could not be

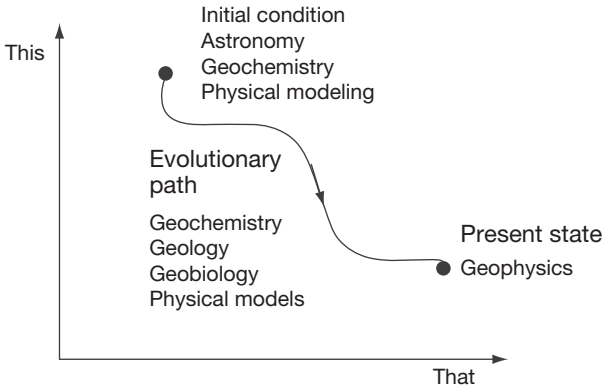


Figure 1 Conceptual view of Earth evolution, identifying the three crucial elements (the initial condition, the evolution path, and the present state) and the sciences that contribute to their understanding. The axes are unimportant, since the diagram is merely a 2-D slice of a multidimensional phase space. They might represent temperature or composition, for example.

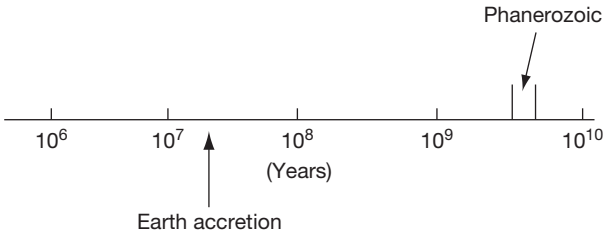


Figure 2 The logarithmic representation of geologic time. The energy budget and rapidity of processes at early times motivate this perspective. A similar view is often taken of cosmology.

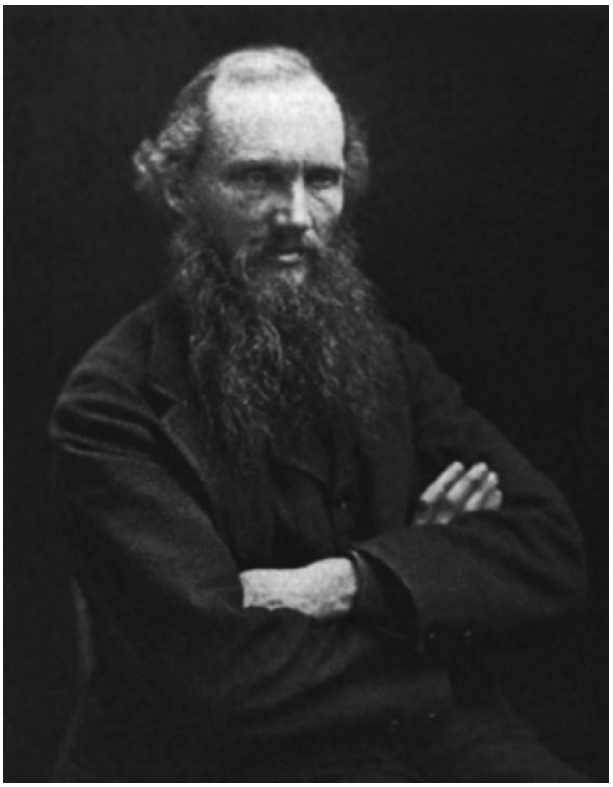


Figure 3 Lord Kelvin played a major role in ideas about Earth evolution and the age of the Earth. Although he got the wrong answer for the Earth's age, he could have obtained roughly the right age had he used for the Earth the method that he used for the Sun.

Table 1 Some important timescales

Process	*Timescale*	*Comments*
Formation of the Earth	10^7–10^8 years	Infrequent large impacts; background flux of small impacts
Cooling of the Earth after a giant impact	1000 years (deepest part of magma ocean) 100–1000 years (condensation of a silicate vapor atmosphere) $\sim 10^6$ years (condensation of steam atmosphere, assuming no additional energy input)	A wide range of timescales, some of which are very fast but nonetheless important. Formation of a water ocean can be fast
Elimination of heat in excess of the thermal state that allows convective heat transport in equilibrium with radioactive heat production	$\sim 10^8$ years	The Earth loses most of the thermal memory of a possible very hot beginning
Decline of impact flux	1×10^8 to 7×10^8 years	The late heavy bombardment at 3.8 Ga may have been a spike rather than part of a tail in the impact flux
Current timescale to cool the mantle by 500 K	5×10^9 to 10^{10} years	Very slow because of the high mantle viscosity

sufficient to explain the features we see. Kelvin's confidence was bolstered by the similar estimate he obtained for the age of the Sun. Indeed, astrophysicists refer to the 'Kelvin time' as a characteristic cooling timescale for a body, defined as the heat content divided by luminosity. We now know that Kelvin was wrong about the Sun because he was unaware of the additional (and dominant) energy source provided by fusion of hydrogen to helium. Ironically, Kelvin could have obtained a correct order of magnitude estimate for the Earth's age had he evaluated the Earth's Kelvin time. For a plausible estimate of mean internal temperature of ~2000 K (a number that would have seemed perfectly reasonable to Kelvin), a heat capacity of 700 J kg^{-1}, a mass of 6×10^{24} kg, and an energy output of 4×10^{13} W, he could have obtained

$$\tau_{\text{Kelvin}} \approx \frac{(6 \times 10^{24}) \times (700) \times (2000)}{4 \times 10^{13}} \approx 2 \times 10^{17} s \approx 7\,\text{Ga} \quad [1]$$

It should not have been unreasonable for him to suppose that this was physically sensible since at that time, the fluidity of the Earth's interior was still in doubt and the concept of efficient convective transport already existed. The subsequent discovery of fission and long-lived radioactive heat sources was *not* the reason he got the wrong answer. Indeed, we now think that the Earth could have a significant part of its heat outflow and dynamics even if those radioactive heat sources did not exist.

Plate tectonics and mantle convection are a central theme as the primary controlling principle of most of Earth evolution. Mobility of the Earth was proposed by Wegener (1912), and the connection to deep-seated motions was also suggested long ago, for example, Bull (1921). The acceptance of these ideas was delayed, especially in the geophysical community, by the perceived absence of compelling evidence together with doubts about physical process: could rocks flow in the way that was needed? So much has been written on this that any attempt to summarize briefly here would be superfluous. However, one aspect deserves comment: Our current view of Earth evolution is not merely the physical picture of how the Earth loses heat and thereby drives the plates; it is the profound interconnection of this to all the other aspects of Earth: the nature of the atmosphere, the existence and persistence of the hydrosphere, the maintenance of the magnetic field, the evolution of the continents, and the evolution of life. Most importantly, these are coupled systems, not one where the mantle is dictating all else. For example, life influences the sediments on Earth, which in turn influence what is cycled back into the mantle, which in turn influences volcanism and pate motions. These interconnections are illustrated in **Figure 4**.

There are some other themes that are important and yet sometimes escape critical attention. The first is the central role of *common processes* rather than *special processes*. In the early, perhaps more speculative, days of scientific theory building, it was acceptable to invoke a special process for planet formation, including the Earth, one that is rare for stars in general. Beginning in the 1950s, the Soviet school of cosmogony developed the view that planet formation is a natural process, building on much older Laplacian ideas of a nebula around the Sun but placing them in the context of a universal concept of a disk around a forming star. Safronov (1972) played a central role in this development. Contemporaneously, the

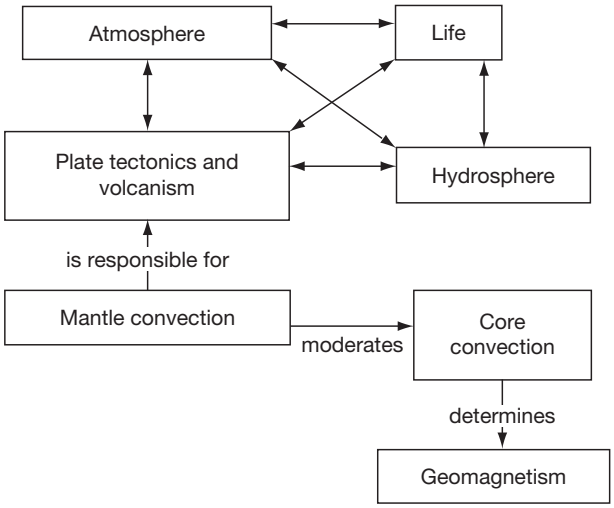


Figure 4 The interconnected aspects of Earth evolution.

astronomical community developed a star formation scenario that naturally developed an attendant disk designed to handle the angular momentum budget of the originating interstellar cloud of gas and dust. Support for this picture grew from the 1980s onward because of astronomical observations. We now have abundant evidence for planets around Sun-like stars, and it is now clear that many of these systems possess terrestrial planets. The prevailing view now is that there is nothing very special about how the Earth came to be though there are many different environments in which Earth-like planets end up. We are reluctant to suggest something very special about how the Earth evolved, even though there is clearly a major difference between the nature of the Earth and the nature of Venus, the most Earth-like of our neighbors in size, though not the most Earth-like in habitability.

Another important theme comes from *meteoritics*. It is widely accepted, yet not entirely obvious, that meteorites inform us about terrestrial planets and about the building blocks for the Earth. Certainly, we have a remarkable amount of information about meteorites: their composition, the timing of their formation, and the conditions that they encountered (both physical and chemical). **Chapter 9.02** by Halliday and Wood provides much information on this. We must ask, however, whether the validity of this approach is as self-evident as it is sometimes portrayed or whether it is more an example of using what we have, as in the classic story of the drunk looking for his lost car keys under the lamppost because it is the only place where he can see. The relationship of materials in the Earth-forming zone to those that formed asteroids (meteorite parent bodies) remains poorly understood.

Also central to current ideas is the role of large impacts rather than dust or merely small bodies in the accumulation of the Earth. Unlike the other themes listed here, this one is more theoretical, though it is consistent with the 'clearing' of dusty disks around stars, taken to infer the formation of planetesimals (though not necessarily requiring the formation of mostly large bodies). This theme originated in work in the 1970s (Cameron and Ward, 1976; Hartmann and Davis, 1975) not only motivated in large part by ideas of lunar formation but also consistent with the Safronov model of

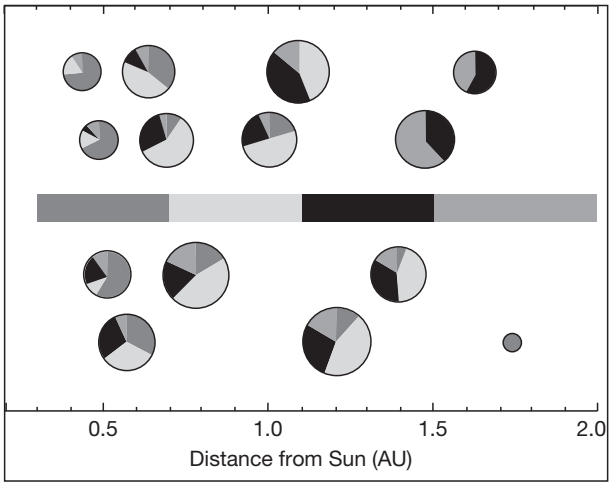


Figure 5 The four outcomes for the final stages of accretion of the terrestrial planets. The shaded bar graph represents an initial distribution of some tracer (e.g., oxygen isotopes or some compositional variable). The circles and their relative size give the final outcome in number spacing and relative masses of the resulting planets. The pie diagrams within each circle represent the relative amounts of material arising from each part of the initial distribution. This shows how substantial mixing is typical; that is, the material that ends up in the Earth is not typically the material that was originally at 1 AU from the Sun. Reproduced from Chambers JE (2004) Planetary accretion in the inner solar system. *Earth and Planetary Science Letters* 223: 241–252.

planet formation as developed further by Wetherill in particular (Wetherill, 1976). Our current ideas of planet formation (e.g., Chambers, 2004; Morbidelli et al, 2012) retain the feature, first noted by Wetherill, that the material that goes into making the Earth comes from a wide range of heliocentric distances but that there is nonetheless some expected variation in final composition simply because of the small number of large bodies that participate. This important idea is illustrated in Figure 5. It is not known whether this is in fact true; the test lies in a better understanding of the compositions (including isotopic makeup) of the other terrestrial planets.

9.01.2 Physical and Chemical Constraints

9.01.2.1 Important Ideas

We cannot figure out the origin and early evolution except to the extent that the Earth has a *memory*. The most obvious memory of a planet is its total mass: we do not know of any way of significantly modifying this after the planet forms. In respect of major elements, the planet is a closed system. The application of cosmochemistry and meteoritics to planets is heavily dependent on this idea of closed systems, especially when we speak of *provenance* (the reservoir of material that was available for forming a planet). It is less obvious but true that a planet will not change its orbital radius significantly. Some other physical attributes such as spin or obliquity or orbital eccentricity can vary significantly (they are not conserved quantities). Aspects of the history of the Earth's rotation are in Chapter 9.09 by Peltier.

The retention of a memory is not guaranteed, especially for physical attributes, even when conservation principles apply. For example, the properties of an ice cube are independent of whether the water molecules were once in the vapor phase or once in a liquid form, at least if the material is in strict thermodynamic equilibrium. The water molecules have no memory. More subtle analysis would however discern whether the isotopic mix of the oxygen on the ice cube was similar or the same as some other ice cubes. One could in this way decide whether the ice cube came from Mars or Earth.

Thermal memories can be elusive. For example, consider a body that cools according to this equation

$$\frac{dT}{dt} = -kT^{a+1} \qquad [2]$$

where T is temperature, t is time, and k is a constant. The exponent a is assumed to be greater than unity, and in many realistic cases (e.g., mantle convection), it may be substantially larger than unity. The solution

$$T(t) = [akt + T(0)^{-a}]^{-1/a} \approx (akt)^{-1/a} \qquad [3]$$

retains poor memory of its initial condition if $a > 1$ and substantial cooling has taken place, that is, $T(t)$ is substantially smaller than $T(0)$. This is a common situation in planets. The loss of memory can be even more striking if there is a long-lived heat source (radioactivity). By contrast, a planet that differentiates (e.g., into the core and mantle) may retain some memory of initial thermal state should the deeper reservoir (e.g., the core) be considerably hotter than the near-surface reservoir. This would appear to be the case for the Earth's core.

Planets are not in thermodynamic equilibrium; that is, one part of the planet is often prevented (by finite diffusivity) from equilibrating with other parts, especially if at least one of the reservoirs is solid. This is a central tenet of many aspects of geochemistry. For this reason, chemical reservoirs are a very important source of memory.

9.01.2.2 Some Useful Estimates

It is useful to have an appreciation of various energy budgets, timescales, and dimensionless numbers. Among terrestrial planets, solar energy is large compared with the energy normally available from the interior of a planet. For example, the incident sunlight on Earth exceeds the current heat flow from the Earth's interior by a factor of 5000. This means that the thermal state of a planet surface is determined by the Sun. However, nonsolar energy (e.g., accretion) can be comparable or more important for short periods of time during the planet formation phase.

The total time of planet accumulation can be long compared with free fall times because the starting material is so widely dispersed. This initial dilution is due in turn to the requirements of angular momentum conservation. For example, suppose we allowed the mass of Earth to be initially dispersed in a volume that is of order one-tenth of a cubic astronomical unit, that is, 1 AU^3 (where 1 AU is the Earth–Sun separation). Then, if an embryo Earth were plowing through this material with a relative velocity of 1 km s^{-1} (a significant fraction of orbital velocity), it would grow at a rate given by

$$4\pi\rho_o R^2 \frac{dR}{dt} = \pi R^2 v \rho \quad [4]$$

where ρ_o is the density of solid matter, $\rho \sim 10 M_\oplus/(1\ \text{AU})^3$ is 12 orders of magnitude smaller, $v \sim 1\ \text{km s}^{-1}$, R is the radius, and a geometric cross section is assumed. This predicts $dR/dt \sim$ few cm per year and a planet formation time of order 100 Ma. Notice that this is independent of whether the accreting material is in the form of dust or larger bodies.

The formation of a planet takes dispersed matter and aggregates it into a mass M of radius R. As a consequence, the gravitational energy changes from a small value to about $-GM^2/R$. By the first law of thermodynamics, this energy release must go somewhere. A small part at most is consumed breaking up material, but most of this energy is converted into heat. If all that energy were stored internally, then the temperature rise is

$$\Delta T \sim \frac{GM}{RC_p} \sim 40000 \left(\frac{R}{R_\oplus}\right)^2 \quad [5]$$

where C_p is the specific heat of terrestrial materials and $R_\oplus$ is the Earth's radius. For the larger bodies, this is so large that some of the rock and iron are converted into vapor; indeed $GM/RL_v \sim 1$ at Earth mass, where L_v is the latent heat of vaporization.

If a planet is initially undifferentiated and then separates into the core and mantle, then the resulting energy released as heat (from the resulting larger but more negative total gravitational energy) is evidently some fraction of this amount. For a core that is one-third of the mass but twice the density of rock and settles an average of one-half the planetary radius, the resulting temperature increase is reduced from eqn [5] by a factor of roughly ten, implying a large effect (thousands of degrees potentially) for the Earth. As with accretional heating, this energy may not be uniformly distributed internally, but unlike that in eqn [5], it is not readily radiated to space.

Suppose we were able to radiate away all of the energy of accretion in time t at a temperature of T_r (ignoring the effect of the Sun or local nebula environment) as IR black body radiation. Accordingly,

$$4\pi R^2 \sigma T_r^4 t \sim \frac{GM^2}{R} \Rightarrow T_r \sim 350 \left(\frac{10^7\ \text{years}}{t}\right)^{1/4} \left(\frac{R}{R_\oplus}\right)^{3/4} \quad [6]$$

This equation requires careful interpretation. If it were indeed true that the Earth formed from small particles (e.g., dust) over ten or a hundred million years, then they would not get hot – the energy of accretion is similar to delivered sunlight (at current solar luminosity) over that period. However, the delivery of mass is highly nonuniform, and much of the mass is delivered in giant impacts. This same equation would say that if you had to radiate away one-tenth of the Earth's energy of formation (the energy arising from impact with a Mars-sized impacting body) at $T \sim 2000$ K appropriate to a silicate vapor atmosphere, then it would take ~ 1000 years. This is far longer than the delivery time of the energy in a giant impact ($\sim$a day or less) implying that very high temperatures are unavoidable.

The early state of the planet depends on the partitioning of energy input between surficial (available for prompt radiative loss) and deep-seated (available for storage and only eliminated if there is efficient heat transfer from the interior). A crude but useful way to estimate this relies on the introduction of a parameter f, the fraction of input energy that is delivered to the deep interior. In the simple case of little energy transport from interior to surface, eqn [5] might then be replaced by

$$\Delta T \approx \frac{fGM(t)}{R(t)} \quad [7]$$

where ΔT is the temperature difference between the near-surface interior and surface at the time when the planet has mass $M(t)$ and radius $R(t)$. It is common practice to think of $f \sim 0.1$ or even less, but the physical basis for this choice is unclear since the heat loss will be high if eqn [7] predicts a high temperature. In other words, f is a function, not a number. However, the prevailing view of the initial state of the Earth is that it is set by a giant impact, presumably the impact that made our Moon. For this view, eqn [7] is not useful, and one must instead appeal to the outcome of impact simulations, for example, Canup (2004). These show that temperature increases of many thousands of degrees are likely, though the temperature distribution is very heterogeneous. This is discussed further in **Chapter 9.03** by Rubie et al. and also figures prominently in the initial condition for the discussion of the resulting magma ocean in the chapter by Solomatov (**Chapter 9.04**).

An enormous range of heat fluxes F from the Earth's interior is possible, both in the accretion epoch and subsequently: in fully molten medium, convective transport can be enormous if the medium is even only slightly superadiabatic. For a superadiabatic temperature difference δT, convective length scale L, and coefficient of thermal expansion α, the mixing length theory (see **Chapter 9.04**) predicts $F \sim \rho C_p \delta T (g\alpha\delta T L)^{1/2}$ assuming that viscosity is small enough to be unimportant. For layer of thickness L, the time to cool by ΔT is accordingly

$$\tau_{cool} \sim 10 \left(\frac{1000}{\Delta T}\right)^{1/2} \left(\frac{\Delta T}{\delta T}\right)^{3/2} \left(\frac{L}{g}\right)^{1/2} \quad [8]$$

Since $(L/g)^{1/2}$ is a short timescale (e.g., hours), we see immediately that the cooling time is short, even for very small temperature fluctuations driving the convection (e.g., of order one degree).

By contrast, cooling times in a system that is mostly solid (and therefore very viscous) can be enormously longer. The corresponding formula for heat flux in this case is $F \sim 0.1 \rho C_p \delta T (g\alpha\delta T \kappa^2/v)^{1/3}$ where κ is the thermal diffusivity and v is the kinematic viscosity. For the usually realistic choice of $\kappa \sim 0.01\ \text{cm}^2\ \text{s}^{-1}$, this predicts a cooling time

$$\tau_{cool} \sim 10^{10}\ \text{years} \left(\frac{L}{1000\ \text{km}}\right) \left(\frac{\Delta T}{1000}\right) \left(\frac{100}{\delta T}\right)^{4/3} \left(\frac{v}{10^{20}\ \text{cm}^2\ \text{s}^{-1}}\right)^{1/3} \quad [9]$$

where the result has been scaled to a choice of viscosity that is roughly like that expected for silicate material near its melting point. The precise value is not important because the real significance of this result lies not in the specific numbers predicted but in the profound difference between cooling of a liquid and cooling of a solid. A planet that melts can lose heat prodigiously but when the Earth is mostly solid, the cooling rate is far slower. One consequence of this is that the time that elapses between the very hot Earth and the formation of an

water ocean can be small. Evidence of an early ocean is discussed in Chapter 9.05 by Karato and also figures in the discussion by Stein and Ben-Avraham on the origin and evolution of continents. It may also be relevant to the initiation of plate tectonics (see Chapter 9.06).

We note also that

$$\frac{GM^2}{R\int_0^\infty Q_{\text{radio}}(t)\mathrm{d}t} \sim 10\left(\frac{R}{R_\oplus}\right)^2 \qquad [10]$$

where Q_{radio} is the chondritic radiogenic heat production, *excluding* short-lived sources (^{26}Al and ^{60}Fe). This emphasizes the importance of gravity as setting the stage for subsequent evolution on Earth.

Work done breaking up materials is small because the strength of materials is small compared to GM^2/R^4, the typical energy per unit volume associated with gravity. On the other hand, $GM^2/R^4K \sim (R/R_\oplus)^2$ where K is the bulk modulus. This means that for Earth-mass planets, gravitational self-compression leads to a higher density (smaller radius) than a body composed of the same constituents but zero internal pressure. This effect is enough to affect significantly the mineral assemblage within larger terrestrial planets, thereby affecting melting behavior, differentiation, core properties, and possible mantle layering. It also means that the Earth's interior can be heated by adiabatic compression alone.

The biggest effect of adiabatic compression is in a possible massive atmosphere, since gas is highly compressible. As a consequence, the radiative temperature of a planet can be much less than the surface temperature of the planet even when the atmospheric mass is a small fraction of the total mass. This assumes that it is opaque (e.g., as in the greenhouse effect, but the argument provided here is not limited to that effect). For example, an adiabatic atmosphere that radiates at a pressure P_e at temperature T_e and has a basal pressure $P_s \gg P_e$ and a mass per unit area $\sim P_s/g$ will have a surface temperature T_s given by

$$T_s \sim T_e\left[10^6\left(\frac{R}{R_\oplus}\right)^2\left(\frac{1\,\text{bar}}{P_e}\right)\left(\frac{M_{\text{atm}}}{M}\right)\right]^\gamma \qquad [11]$$

where $T \propto P^\gamma$ is the adiabatic relationship and M_{atm} is the atmospheric mass (assumed small). The factor of 10^6 demonstrates the remarkable blanketing effect that is possible even for an atmospheric mass that is less than the planet mass by a factor of a thousand.

Planetary embryos form early enough that they are imbedded in the solar nebula. Gravitational attraction increases the nebula gas density near the embryo surface. For an isothermal atmosphere of negligible mass,

$$\frac{\rho(r=R)}{\rho(r\to\infty)} = \exp\left[\frac{GM}{Rc^2}\right] \qquad [12]$$

where c is the speed of sound for the nebula (primarily hydrogen). Since the nebula is very low density, this is only of interest for surface conditions (e.g., ingassing of nebula at the surface of a magma ocean) if GM/Rc^2 is larger than $\sim$5 or 10. For a Mars mass and $T \sim 300$ K, $GM/Rc^2 \sim 12$. For such an embryo, the nebula might be $\rho \sim 10^{-9}$ g cm^{-3} and the surface density could then be $\rho \sim 10^{-4}$ g cm^{-3}, potentially optically thick, and with a surface pressure ($\sim$1 bar with a large uncertainty) that allows modest ingassing. Thus, embryos larger than Mars, including the growing Earth, could have had a massive atmosphere of near-solar composition. However, the planetary evidence suggests that this effect is small. For example, the neon to argon ratio for the Earth is much less than the solar nebula ratio even though the ingassing ability (predicted by Henry's law) is similar. This is discussed further in Chapter 9.05 by Karato. Presumably, the nebula was eliminated early in the period of growth of large embryos.

9.01.3 Commentary on Formation Models

Astronomical observations of newly forming stars indicate the presence of disks of gas and dust. The material in the disk is typically a few percent of a solar mass, more than sufficient to explain the observed planetary mass of our system or other systems discovered thus far. If the disk has solar composition, then the amount of condensable material (as rock or iron) internal to a few AU is sufficient to explain the masses of the terrestrial planets. However, the disks extend to tens of astronomical units or more, and this is enforced by the angular momentum budget of the originating interstellar cloud. This angular momentum not only is responsible for the possibility of planet formation but also guarantees that the terrestrial zone is a small part of a much bigger picture. This means that we probably cannot understand formation of terrestrial planets without some understanding of the formation of the gas giants, especially Jupiter. The disks have a radial variation in temperature, wholly or partly because of the energy release of the central body (the forming star). As a consequence, the region within a few AU of the star is hundreds of K or more, sufficient to avoid the condensation of water ice. The absence of large amounts of water in the terrestrial zone in our solar system is interpreted as evidence for terrestrial planet formation internal to the 'snow line' (the place outward of which water ice can condense). Typically, this temperature is of order 160 K (low compared with 270 K because of the very low vapor pressures characteristic of such a nebula). Chapter 9.05 by Karato discusses the source of water on Earth, which is presumably external to the terrestrial planetary zone.

The central ideas in current models of terrestrial planet accretion are three: planetesimals formation, runaway and oligarchic growth, and late-stage accumulation. Planetesimal formation refers to the accumulation of bodies of order kilometers in size. Runaway and oligarchic growth are two dynamic stages of a process that we can (for our purpose) lump into the rapid accumulation planetary embryos up to of order Mars in mass but in closely spaced low-eccentricity and inclination orbits. The last and slowest stage proceeds for hundreds of Moon- and Mars-sized objects to the terrestrial planets that we see. According to astronomical observations, this last stage is probably occurring after the solar nebula has been removed (by accretion onto the Sun or expulsion to the interstellar medium). Chapter 9.02 by Halliday and Wood connects this physical picture to the cosmochemical evidence.

Planetesimal formation remains mysterious even though there is no doubt that it occurred, since its consequences is expressed among the meteorites arising from differentiated

asteroids that must have formed in the first million years after the formation of the solar nebula. They may have formed by the poorly understood process of sticking of dust particles during very slow velocity collisions in the dusty gas. They may alternatively have formed by a gravitationally mediated process as the dust settled toward the midplane of the disk. Gravitational instabilities are an attractive mechanism, especially given the uncertainties of the physics of sticking, but the relevant fluid dynamics is still debated (Weidenschilling, 2006). This remains one of the major puzzles of planet formation.

The next stage, from planetesimals to Moon and Mars mass embryos, is perhaps better understood theoretically but less well founded in direct observation. This stage, lasting 10^5–10^6 years, is rapid because of gravitational focusing. Bodies encounter each other at velocities considerably less than the escape velocity from the larger body, and as a consequence, the cross section for collision is far larger than the geometric cross section, perhaps by as much as a factor of a thousand. The planetesimal swarm is 'cold' (i.e., the random velocities of the bodies are very small compared with Keplerian orbital velocities). This process of embryo growth is terminated by isolation: when the orbits are so nearly circular, the bodies reach a stage where there are no crossing orbits (i.e., no overlap of their gravitational spheres of influence). From the point of view of understanding the bodies that we see, this stage is of great interest in three ways: (1) These embryos form quickly and therefore may be hot, both because of possible short-lived radioactive elements and because of accretional heating (cf. the scaling discussed in **Section 9.01.2**). They could therefore form cores and primordial crusts. In this sense, the primordial crusts and current cores of the terrestrial planets have the potential to predate the formation of the planets. (2) The largest of these embryos could conceivably be a surviving planet, most plausibly Mars. In this picture, Mars is special as an isolated outlier in the formation process. (3) Embryos form while the solar nebula is still present and could therefore have a component of solar composition gas. Equation [12] suggests that this is small however.

The late-stage aggregation is conceptually like that envisaged long ago in the work of Safronov (1972), whereby the scattering of bodies causes growth of eccentricity and inclination allowing the orbits to cross. It has received a lot of attention in recent times because of the development of N-body codes that can handle the outcome of a population of bodies scattering from one another gravitationally and occasionally colliding (Chambers, 2004). The main shortcoming of these calculations is the failure to take full account of what actually happens when two bodies have a very close encounter (a quite common occurrence) or collide. Tidal disruption is possible in a close encounter (i.e., the outcome is not necessarily the two intact bodies that existed before encounter), and collisions do not always lead to a clean merging of the two bodies. Close encounters can also create additional bodies through tidal disruption.

It is easy to see by order of magnitude (a slightly more sophisticated version of eqn [4]) that this late-stage process can take as long as 100 Ma. However, it is stochastic and it involves large bodies. As a consequence, it is not possible (and may perhaps never be possible) to say exactly what sequence of major events took place in the formation of a particular planet.

9.01.4 Commentary on Early Evolution Models

The high-energy events of late-stage accumulation are expected to play a central role in setting the stage for Earth structure and evolution. The building blocks are a set of planetary embryos that almost certainly form iron-rich cores because of the combined effects of gravitational energy release in a relatively short time and the possible effects of ^{26}Al heating. This early but important differentiation event is characterized by relatively modest temperatures (~2000 K or less) and pressures (20 GPa or less, perhaps a lot less) appropriate to bodies in this size range. It is likely and perhaps significant for some constituents (e.g., noble gases) that this phase occurs in the presence of the solar nebula. In highly idealized numerical models (Chambers, 2004; Morbidelli et al, 2012), it is usually assumed that nearly all of the Earth's mass accumulates in the subsequent merging of these massive embryos. In reality, it is not known how much material arrives in small bodies (planetesimals), whose origin could be either the bodies that were not swept up during runaway or debris created during frequent close encounters in the final orbit-crossing phase. Giant impacts will certainly create extensive melting and at least a transient magma ocean, especially if one assumes that the interiors of the embryos prior to impact are at or even above the solidus of mantle minerals. This is a reasonable assumption because of the limited time available for elimination of earlier heating, together with the higher radioactive heating of that epoch. It is also possible that there is a persistent magma ocean, even in the long intervals of time (~10^7 years) between the giant impacts, sheltered by a steam atmosphere greenhouse that is sustained by delivery of a 'rain' of small bodies that heat its base. This kind of magma ocean is less certain than the transient high-energy ocean that is present immediately after a giant impact. These magma ocean scenarios not only are analyzed in detail in **Chapter 9.04** but also figure prominently in the discussion of core formation in **Chapter 9.03** by Rubie et al.

Of course, short-lived high-energy events could have a bigger role than sustained lower-energy events in determining the composition of the current core and mantle because they can create extensive melting in which core–mantle separation is especially efficient. There are two central questions that arise in this picture:

(1) To what extent is the Earth's core formation accomplished by merging the cores of embryos rather than by separation of iron alloy from the mantle within the Earth?
(2) To what extent is the separation of the core from the mantle and possible internal differentiation of the mantle accomplished in a magma ocean environment?

These issues figure prominently in **Chapter 9.03** by Rubie et al.

In simulations designed to understand the origin of the Earth's moon, smoothed-particle hydrodynamics (SPH) is most commonly used to describe the outcome of an impact involving a Mars-sized projectile and a mostly formed Earth (Canup, 2004). This represents perhaps the last (and perhaps most important) giant impact. The core and mantle are tracked, and one can follow the extent to which these constituent parts of projectile and target 'mix.' The outcome is

essentially stochastic; that is, it varies considerably as a function of details such as impact parameter and mass ratio of projectile to target, parameters that we can never hope to establish deterministically. However, in typical simulations, only around 10% of the projectile (and thus at most a few percent of an Earth mass) is placed in orbit and available for making the Moon. Very little of the Earth is placed into orbit. There have been some recent suggested alternatives (see Canup (2013)) but these remain controversial. In any event, nearly all of the impactor either merges immediately with the Earth or crashes back into the Earth within several hours of the initial impact. The iron core of the projectile may be stretched into filaments or broken into blobs. These simulations are low resolution, so even the finest scale features of SPH are hundreds of kilometers in size. It seems likely that much of the iron is emulsified to small scales and equilibrates with the mantle, but this is not certain.

The transitional phase between a mostly molten mantle and an almost entirely solid mantle is still poorly understood. In respect of thermal history, it could be argued that the details of this phase are unimportant. In respect of the geochemical evolution, it is probably very important for setting up the conditions for formation or preservation of the earliest crust (**Chapter 9.02**).

The initially hot core will be allowed to cool by the overlying mantle and will therefore convect, allowing for the possibility of magnetic field generation. Nimmo deals with core evolution in **Chapter 9.08**, particularly the puzzle of whether there is enough energy in core cooling and inner growth to explain the energy budget demanded by the persistence of a dynamo throughout nearly all of geologic time. The answer to this puzzle remains uncertain.

What about the origin of plate tectonics? On the one hand, this could be viewed as a geologic question, provided that there was agreement about the distinctive signatures of plate tectonics in the rock record. On the other hand, this could be viewed as a fluid dynamic question (albeit one that is unavoidably coupled to the crustal evolution and rheological puzzle of a fragmenting lithosphere). Sleep discusses both aspects of this unresolved question.

Finally, we come to life. The origin of life on Earth remains one of the central scientific problems of our time, even though many in the scientific community have adopted the view that this origin was 'easy,' by which they mean that the physical conditions, amounts of material, energy budget, and abundance of liquid water and surfaces allow for conditions that are suitable. Less controversial but perhaps as important, it is abundantly clear that biological evolution has affected the physical aspects of Earth evolution profoundly, just as the physical conditions have mediated biological evolution. In many ways, this geobiological question is the most exciting frontier of Earth science, along with the more immediate issues of climate change (Earth evolution on a much smaller timescale).

9.01.5 Outstanding Questions

What are the main issues that arise in our understanding of Earth origin and evolution? I end by providing in the succeeding text a list of ten questions, linked to the subsequent chapters in this volume. In some cases, these emerge naturally from the perspectives of the various authors, while in other cases, they show the interconnections that might be less evident from a single chapter.

1. *What is the legacy of Earth formation in the planet as we see it?* What range of accretion scenarios is consistent with Earth? What is the role of chance? What is the role of distance from the Sun? How does geochronology compare with physical estimates of formation times? (Halliday and Wood)
2. *How is core formation in Earth expressed in the composition of the Earth now and the thermal structure and evolution?* Can we figure out the core formation conditions from the abundances of siderophiles in the mantle? Or will we find that the pattern is a mix of low-temperature, low-pressure and high-temperature, high-pressure events? (Rubie et al.)
3. *What was the duration of the early magma ocean phase of the Earth and what memory (if any) is there of this phase in the early mantle differentiation or crust formation?* Is there an opportunity for melt–solid separation in the early mantle and can this be reconciled with current models of mantle structure and composition? What of the role of an early dense atmosphere? How long did it take for a magma layer within Earth (e.g., in the transition zone) to completely freeze? (Solomatov)
4. *Where did the Earth's water come from? How much do we have and how much of it has cycled between the mantle and hydrosphere over geologic time?* Did water play an essential role in the onset of plate tectonics? (Karato)
5. *Why does Earth have plate tectonics?* Mantle convection is easily understood; plate tectonics still resists understanding because of our imperfect knowledge of how the lithosphere fails. The perspective of geologic time helps us solve this puzzle. (Sleep)
6. *Why and when did continents evolve?* Are continents a sideshow in the overall dynamics of Earth evolution (as passengers on plates) or do they play a central role in how the system evolves? (Stein and Ben-Avraham)
7. *How did the Earth's mantle temperature evolve through time? Can we reconcile simple convective models with the radiogenic heat supply, the geologic evidence, and our understanding of plate tectonics?* The mismatch between known heat generation from long-lived radioactivity and the actual heat flow out of the Earth persists and is incompletely resolved in simple cooling models, suggesting that the evolution is more complex.
8. *Can we reconcile Earth thermal evolution with the persistence of the geomagnetic field throughout geologic time? When did the inner core form?* The mantle governs the rate at which the core cools, and models acceptable on the mantle side suggest a possible problem for core heat flows, including the possibility that the inner core was absent for a major part of early Earth history. But if so, can we supply enough energy to run the Earth's magnetic field? (Nimmo)
9. *What do variations in Earth rotation tell us about the nature of the Earth and its evolution?* Have we made sufficient use of rotation history, including true polar wander, in our attempts to reconstruct how the Earth evolved? (Peltier)

10. *How are the geophysical and biological evolutions of the Earth interrelated?* What does the history of life tell us about plate tectonics and the history of water on Earth? To what extent is the coevolution of life and Earth's physical attributes so strongly coupled that we connect major events in biological evolution to major events in geologic history? (Retallack).

References

Bull AJ (1921) A hypothesis of mountain building. *Geological Magazine* 58: 364–367.

Burchfield JD (1975) *Lord Kelvin and the Age of the Earth*, p. 260. New York: Science History Publications.

Cameron AGW and Ward WR (1976) The origin of the Moon (abstract). In: *Lunar Science VII*, pp. 120–122. Houston: Lunar Science Institute.

Canup R (2013) Lunar conspiracies. *Nature* 504: 27–29.

Canup RM (2004) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.

Chambers JE (2004) Planetary accretion in the inner solar system. *Earth and Planetary Science Letters* 223: 241–252.

Hartmann WK and Davis DR (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24: 504–515.

Hopkins W (1839) Preliminary observations on the refrigeration of the globe. *Philosophical Transactions of the Royal Society of London* 129: 381–385.

Morbidelli A, Lunine JI, O'Brien DP, Raymond SN, and Walsh KJ (2012) Building terrestrial planets. *The Annual Review of Earth and Planetary Sciences* 40: 251–275.

Safronov VS (1972) *Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets*, p. 206. Jerusalem: Israel Program for Scientific Translations.

Wegener A (1912) Die Entstehung der Kontinente. *Geologische Rundschau* 3: 276–292.

Weidenschilling SJ (2006) Models of particle layers in the midplane of the solar nebula. *Icarus* 181: 572–586.

Wetherill GW (1976) The role of large bodies in the formation of the earth. In: *Proceedings of the Seventh Lunar Science Conference*, pp. 3245–3257.

This page intentionally left blank

9.02 The Composition and Major Reservoirs of the Earth Around the Time of the Moon-Forming Giant Impact

AN Halliday and BJ Wood, University of Oxford, Oxford, UK

© 2015 Elsevier B.V. All rights reserved.

9.02.1	Introduction	11
9.02.2	Key Features of the Earth and Moon	11
9.02.3	The Birth of the Solar System	12
9.02.4	Meteorites	13
9.02.5	Meteorites and the Composition of the Earth and Its Primary Reservoirs	14
9.02.6	The Circumstellar Disk and the Composition of the Earth	18
9.02.7	Dynamics of Planet Formation	19
9.02.8	The Age of the Earth	21
9.02.9	Short-Lived Nuclides and Early Processes	22
9.02.10	Rates of the Earth's Accretion and Differentiation	24
9.02.11	The Age of the Moon	26
9.02.12	First Principles of Chemical Constraints on Core Formation	27
9.02.13	Explanations for the 'Excess Siderophile Problem'	28
9.02.14	The 'Deep Magma Ocean' Model of Core Formation	28
9.02.15	Core Segregation During the Growth of the Earth	29
9.02.16	Oxidation State of the Earth During and After Accretion	30
9.02.17	Isotopic Evidence for Impact-Induced Losses During the Earth's Accretion	31
9.02.18	Hidden Reservoirs and the Composition of the Earth	35
9.02.19	Concluding Overview	36
Acknowledgments		38
References		38

9.02.1 Introduction

A treatise on geophysics needs some summary explanation of the current thinking on how Earth achieved its current composition and formed its primary reservoirs. This subject has taken enormous strides in the past decades thanks to significant improvements in mass spectrometry in particular. However, additional discoveries are being fuelled by exciting developments in other areas and are leading to joint research at the interfaces between disciplines traditionally viewed as scientifically distinct. The most significant of these are as follows:

- *Accretion dynamics* – theoretical calculations of the behavior of gas and dust in the solar nebula; the construction of planetesimals, planetary embryos, and planets; and the fluid dynamics of primordial mixing and differentiation
- *Cosmochemistry* – the isotopic and chemical history of the solar system deduced from meteorites and lunar samples
- *Mineral physics and experimental petrology* – the simulation experimentally and theoretically of the phases present in the Earth's interior and their physical properties and behavior
- *Observations* – the observations of planets, stars, disks, and exosolar planetary systems

In this chapter, we summarize briefly these lines of evidence and present the current thinking on the formation and primordial differentiation of the Earth–Moon system. We focus on constraints from cosmochemistry and experimental petrology in particular because other chapters provide complementary information. In particular, the introductory chapter by Stevenson (Chapter 9.01) covers aspects of the accretion dynamics as does the chapter by Rubie et al. (Chapter 9.03). There are now observations of Earth-sized planets around other stars, and in the coming years, the search for Earth-*like* objects represents a major thrust in experimental astrophysics. Indeed, the links between Earth sciences and astrophysics have probably never been stronger.

9.02.2 Key Features of the Earth and Moon

Any theory for how the Earth formed and acquired its present composition and distribution of component reservoirs after the Moon-forming giant impact has to explain several things:

1. Rocky terrestrial planets grew from the same disk as the gas giants Jupiter and Saturn although plausibly at different stages in its development.
2. Terrestrial planets and differentiated asteroidal objects are depleted in moderately volatile elements relative to refractory elements.
3. The Moon is larger relative to the size of its host planet than any other moon in the solar system.
4. The Moon carries most of the angular momentum of the Earth–Moon system.
5. The oxygen, silicon, titanium, chromium, and tungsten isotopic compositions of the Earth and Moon are identical

despite their different chemical compositions and accretion histories.
6. The Earth has a larger core proportionally speaking than does the Moon, Mars, or asteroid Vesta.
7. The Earth has a more oxidized mantle than the Moon, Mars, or asteroid Vesta.
8. Highly siderophile (metal-loving) elements in the Earth's mantle are inconsistent with low-pressure core formation.
9. The Earth has significant water and a high H/N ratio.
10. The Earth, unlike the Moon, Mars, or asteroid Vesta, lacks a geologic record for the period prior to 4.0 Ga.
11. Time-integrated parent/daughter ratios determined from isotopic data provide evidence of changes in composition during the accretion history of the protoplanetary material that built the Earth and Moon.

Most of the above features are still a matter of study. To tackle these subjects, this chapter first provides a brief and elementary summary of how the Sun and solar system formed and what we know about terrestrial planet bulk compositions. We then give a synopsis of the range of dynamic and isotopic constraints on the formation of the Earth and Moon, following this with a summary of the huge field of experimental and theoretical petrology that provides insights into how Earth must have changed during its accretion history.

9.02.3 The Birth of the Solar System

The Hubble Space Telescope has provided us with fascinating images of numerous young solar mass stars forming in giant molecular clouds such as the Eagle and Orion Nebulae. It is thought likely that our Sun originated in a similar kind of environment. New evidence confirms a long-held suspicion that the Sun did not form spontaneously but that cloud collapse was triggered by a shock wave originating from a nearby star. Such shock waves might be produced, for example, from supernovas – the explosive finale to the short life of giant stars. This idea, developed by Al Cameron of Harvard (e.g., Cameron and Truran, 1977), among others, appeared to be consistent with the discovery of the daughter products of a range of short-lived nuclides in meteorites (**Table 1**). The half-lives of these nuclides (100 ky–100 Ma) are such that they must have been synthesized in stars shortly before the start of our solar system (**Figure 1**). The most important paper outlining the likely explanation for the synthesis of the various nuclides in stars is still the work of Burbidge et al. (1957). A variety of more recent papers explore the growing body of evidence from short-lived nuclides regarding the later inputs to the solar nebula (Busso et al., 2003; Huss et al., 2009; Meyer and Clayton, 2000; Meyer and Zinner, 2006; Truran and Cameron, 1978; Wasserburg et al., 1994, 1996, 1998). The early Sun may have been associated with relatively energetic irradiation (Feigelson et al., 2002a,b). It has been proposed (Shu et al., 1997) that the conveyor belt of material being swept in toward the Sun may have led to some light short-lived nuclides being generated locally and heterogeneously (Lee et al., 1998). These might then have been scattered across the disk in early-formed condensates by X-winds (Shu et al., 1997). Although local production in the early Sun is feasible for a range of nuclides (Gounelle et al., 2001; Leya et al., 2003) for most, the abundances would be too low to be a match for those found in the meteorites and this X wind model is not well accepted (Desch et al., 2010). Furthermore, some of these nuclides could only have been made in a large star that predated the solar system. They were then introduced to the gas and dust that ultimately became our Sun and planets. It is possible that this explosion actually triggered the collapse of the cloud material onto a core that became the solar nebula and nascent Sun.

Problems have arisen with this theory. First, a supernova is so powerful that rather than causing collapse, it may in fact

Table 1 Short-lived nuclides that have been demonstrated to have existed in the early solar system as a result of anomalies in the abundance of their daughter isotopes

Radionuclide	*Daughter*	*Half-life (million years)*
^{10}Be	^{10}B	1.5
^{26}Al	^{26}Mg	0.71
^{36}Cl	^{36}S, ^{36}Ar	0.30
^{41}Ca	^{41}K	0.10
^{53}Mn	^{53}Cr	3.7
^{60}Fe	^{60}Ni	2.6
^{92}Nb	^{92}Zr	35
^{107}Pd	^{107}Ag	6.5
^{129}I	^{129}Xe	15.7
^{135}Cs	^{135}Ba	2.3
^{146}Sm	^{142}Nd	68
^{182}Hf	^{182}W	8.9
^{205}Pb	^{205}Tl	17
^{244}Pu	^{131,132,134,136}Xe	80
^{247}Cm	^{235}U	16

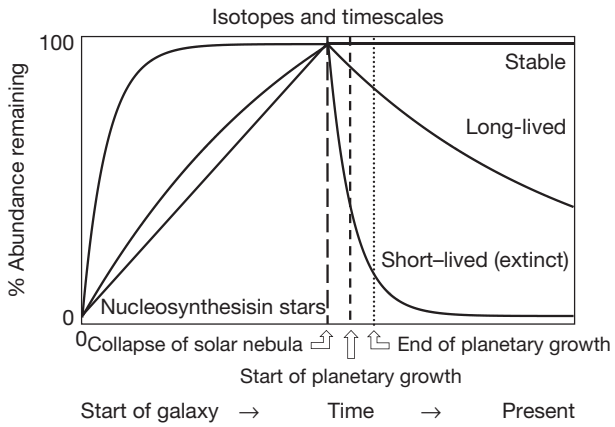


Figure 1 Most solar system nuclides heavier than hydrogen and helium were produced in stars over the history of our galaxy. This schematic figure shows the difference between nuclides that are stable, those that have very long half-lives (such as ^{238}U used for determining the ages of geologic events and the solar system itself), and those that have short half-lives of $<10^8$ years, assuming all were produced at a constant rate through the history of the galaxy. The short-lived nuclides decay very vast and provide crucial insights into the timescales of events, including planet formation, immediately following their incorporation into the solar nebula.

shred a molecular cloud unless located at sufficient distance that much of the energy and material have already dissipated (Foster and Boss, 1996, 1997; Stone and Norman, 1992; Vanhala and Cameron, 1998). The time taken to reach the location of the potential new star then becomes sufficiently long (hundreds of thousands of years) that it would be difficult, though not impossible, to detect the former presence of a nuclide such as ^{41}Ca with a half-life of just 100 ky, because most of it would have decayed.

The second problem is that the relative abundances of many of the short-lived nuclides are not well explained in terms of the production ratios expected in a supernova (Wasserburg et al., 1998). Therefore, attention has turned to other stellar sources, the most likely of which is an AGB (asymptotic giant branch) star. AGB stars are relatively common (<8 solar mass) stars that have been shown spectroscopically to be enriched in some heavy elements such as lead (Pb) supporting the idea that they represent the most likely 'factory' for *s*-process nuclides. This was confirmed by the discovery of technetium (Tc) in the spectrum of an AGB star (Merrill, 1952). With no stable nuclei, Tc can only be present if it is being synthesized. AGB stars shed material into space but are much less energetic than supernovas so could be located close to the site of the Sun's formation. The relative proportions of many of the former short-lived nuclides made better sense with an AGB star model. *r*-process nuclides such as ^{129}I, ^{182}Hf, and ^{244}Pu have to have been produced in a larger star but they possess relatively long half-lives (**Table 1**). The stars that produced them may therefore have been unrelated to the formation of the Sun. The principal problem with the AGB model is ^{60}Fe, which has a half-life of just 1.5 Ma. Although it is not an *r*-process nuclide, ^{60}Fe is difficult to make in large quantities in an AGB star. The AGB star model appeared to be consistent with the low initial solar system abundance of ^{60}Fe (Wasserburg et al., 1998) inferred from the first discoveries of this nuclide by Shukolyukov and Lugmair (1993). The initial abundance of ^{60}Fe is now known to have been higher than previously suspected, however (Mostefaoui et al., 2004). With this new discovery, plus the difficulty of finding a match for the initial abundance of ^{53}Mn, supernova contributions from two kinds of stars ($\leq$11 $M_\odot$ and 12–25 $M_\odot$) are recognized as potential candidates for a trigger (Huss et al., 2009).

The short-lived radionuclides discussed in the preceding text provide additional information relevant to the early solar system. First, they provide a method for determining timescales from isotopic measurements of the daughter isotopes. Second, they provide a source of heat that would have been available for melting the earliest planetesimals. Both of these are very relevant to understanding how Earth formed.

9.02.4 Meteorites

The Sun contains 99.9% of the mass of the solar system and spectroscopic measurements of its chemistry provide an indication of the composition of the circumsolar disk from which the planets grew. However, the compositions of the planets are highly variable and they have been left as a heterogeneous collection of parts of a small fractionated residue after much of the disk had dissipated or been consumed by the Sun. A better set of archives for studying the early solar system and its average composition is found in meteorites – solid extraterrestrial materials that have survived passage through the Earth's atmosphere. These have the added advantage that they can be studied in the lab at much higher precision than can the composition of the Sun. A detailed description of meteorites is beyond the scope of this chapter but two excellent reviews of the main undifferentiated and differentiated meteorite classes can be found elsewhere (Brearley and Jones, 1998; Mittlefehldt et al., 1998). In order to explain how we know certain things about the Earth's composition and origins, it is necessary to provide a brief explanation of this important archive.

The exact origins of meteorites are varied and the subject of some debate. The range and number of examples of each type has expanded significantly with the acquisition of Antarctic and Saharan meteorites. We refer to the body from which a meteorite came as its *parent body*. Meteorites are usually divided into three main types: *chondrites*, *achondrites*, and *irons*.

Chondrites are complicated mixtures of dust, glass and metal thought to represent accumulations of solar system debris that was floating in the circumstellar disk when the Sun formed. The average composition of chondrites is generally similar to that expected from unprocessed circumstellar solids, little affected by melting and loss of volatiles or metal. The most primitive are carbon-rich and are called *carbonaceous chondrites*. *Ordinary chondrites* are noncarbonaceous chondrites containing metal. *Enstatite chondrites* are a less common group of highly reduced chondrites. In some cases, a well-observed meteor trajectory has been linked to the meteorites found on the ground, and from these, it has been established that at least some ordinary chondrites come from the asteroid belt between Mars and Jupiter. Most chondrites contain calcium–aluminum *refractory inclusions* (*CAIs*) enriched in those elements expected to condense at very high temperatures from a hot nebular gas. These may have been ejected from the inner portions of the disk and are the oldest objects identified as having formed in the solar system. CAIs from the Efremovka chondrite have been dated by $^{235/238}$U–$^{207/206}$Pb at 4.5672 ± 0.0006 Ga (Amelin et al., 2002), which is close to all subsequent high-precision attempts to define the canonical start to the solar system (**Figure 2**). Most chondrites also contain *chondrules*, which are spheroidal objects with textures thought to reflect rapid heating, melting, and quenching of preexisting silicate material. Many of these formed as much as 2 million years after the start of the solar system (Bizzarro et al., 2004 2005; Kita et al., 2006; Russell et al., 1996).

Achondrites are silicate-rich mafic and rare ultramafic igneous rocks not too dissimilar from those formed on Earth but with different chemistry and isotopic compositions. They mainly represent basaltic rocks and cumulate from other planets and asteroids. It is possible to group these and distinguish which planet or asteroid they came from. This is done using their oxygen isotopic compositions. Oxygen is extremely heterogeneous isotopically in the solar system, and meteorites with similar chemistry and petrography seem to have specific compositions such that it is thought that they can be linked to parent bodies from different regions (**Figure 3**). This is one way in which *martian* meteorites are grouped although the more specific link to Mars is established through the dominance of young meteorite ages requiring a geologically active body and the similarity in volatile compositions to the martian

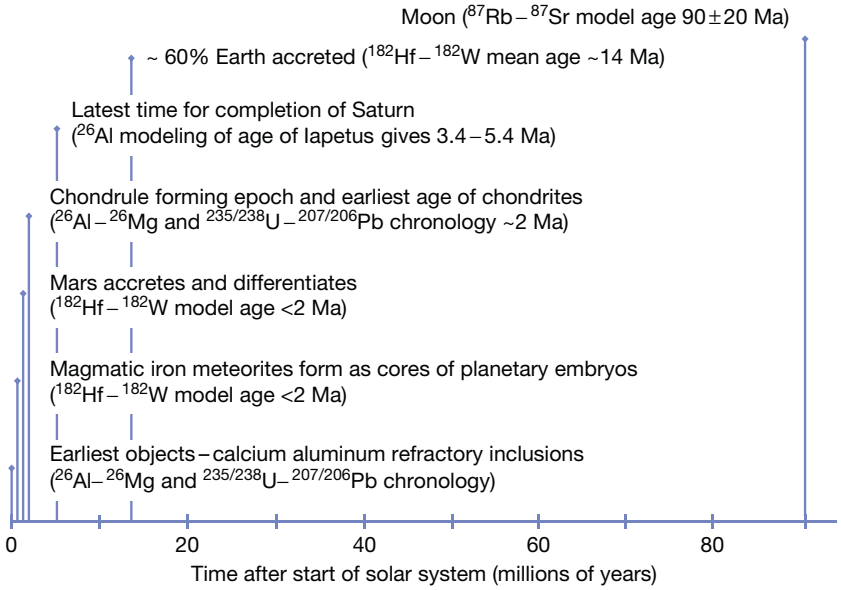

Moon (^{87}Rb–^{87}Sr model age 90±20 Ma)
~ 60% Earth accreted (^{182}Hf–^{182}W mean age ~14 Ma)
Latest time for completion of Saturn
(^{26}Al modeling of age of Iapetus gives 3.4–5.4 Ma)
Chondrule forming epoch and earliest age of chondrites
(^{26}Al–^{26}Mg and $^{235/238}$U–$^{207/206}$Pb chronology ~2 Ma)
Mars accretes and differentiates
(^{182}Hf–^{182}W model age <2 Ma)
Magmatic iron meteorites form as cores of planetary embryos
(^{182}Hf–^{182}W model age <2 Ma)
Earliest objects–calcium aluminum refractory inclusions
(^{26}Al–^{26}Mg and $^{235/238}$U–$^{207/206}$Pb chronology)
0
20
40
60
80
Time after start of solar system (millions of years)

Figure 2 The current best estimates for the timescales over which very early inner solar system objects and the terrestrial planets formed. The approximated mean life of accretion is the time taken to achieve 63% growth at exponentially decreasing rates of growth. Reproduced from Halliday AN and Kleine T (2006) Meteorites and the timing, mechanisms and conditions of terrestrial planet accretion and early differentiation. In: Lauretta D, Leshin L, and MacSween H (eds.) *Meteorites and the Early Solar System II*, pp. 775–801. Tucson, AZ: University of Arizona Press.

atmosphere, as sampled by Viking. The *eucrites*, *howardites*, and *diogenites* are all thought to come from the asteroid 4 Vesta. Other important groups are the *angrites*, which are extremely depleted in volatile elements (Figure 4), and the *ureilites*, which appear to have been produced by melting of a carbonaceous chondrite-like object that has not undergone core formation. Achondrites and lunar samples are immensely important for providing evidence for how differentiated planets other than Earth formed. An important feature of all of these samples of differentiated planetary bodies is that they indicate more extreme volatile depletion than is found in chondrites. The same is true of the Earth and Moon.

Irons are exactly what you would expect – metallic iron (Fe) with large amounts of nickel (Ni) and some blebs of iron sulfide. Most represent the remains of the metallic cores of early-formed small asteroidal bodies. Based on their cooling rates determined from Fe–Ni exsolution features, some ('non-magmatic') types are thought to have formed at shallow depths from impact melting of asteroidal surfaces (Wasson, 1985). Magmatic group irons typically are more volatile-depleted and appear to be derived from the cores of objects of up to hundreds of kilometers in size. They provide evidence on the very earliest stages of core formation in the solar system.

The exposure age of a meteorite to cosmic rays can be determined from the nature of the cosmogenic nuclides it contains. Iron meteorites can have very long exposure ages – up to hundreds of millions of years – attesting to a long history of survival in space. Stony meteorites have much shorter exposure ages. They do not survive well, which explains why we have no examples of silicate mantle material that must have been complementary to the magmatic irons. All of the stony meteorites appear to have been knocked off their parent bodies from shallow depths in the relatively recent past. We do however have some meteorites that contain metal mixed in with silicate that may have formed at some kind of core–mantle boundary in a small object. These are called *pallasites*. *Mesosiderites* are also metal–silicate mixtures but are thought to be the products of impact melting.

9.02.5 Meteorites and the Composition of the Earth and Its Primary Reservoirs

The Earth is depleted in those components that are expected to have partially resided in the gas phase of the solar nebula at low temperatures, such as H, N, O, and all of the noble gases. Although this is obvious, the magnitude of the depletion is in part gauged from comparing the Sun's composition with those of meteorites. The most primitive kind of chondrite is called a CI carbonaceous chondrite. If one compares compositional data for the Sun with those for CI carbonaceous chondrites, the agreement is very good for all but the most volatile elements. Thus, these primitive meteorites provide examples of the materials that went to produce the volatile-depleted terrestrial planets.

The estimated abundances of the elements in the solar system, dominated by the Sun (Figure 5), range through 13 orders of magnitude and are most easily compared by plotting on a log scale such that the number of atoms of Si is 10^6. Hydrogen and helium dominate over all other elements, reflecting production in the big bang and in first-generation stars. Other features reflect the processes of nucleosynthesis in stars, as first explained by Burbidge et al. (1957). The essential features are as follows.

First, the abundances of the elements decrease with increasing atomic number. This is because most of the elements are made from lighter elements. There is an anomalous peak in

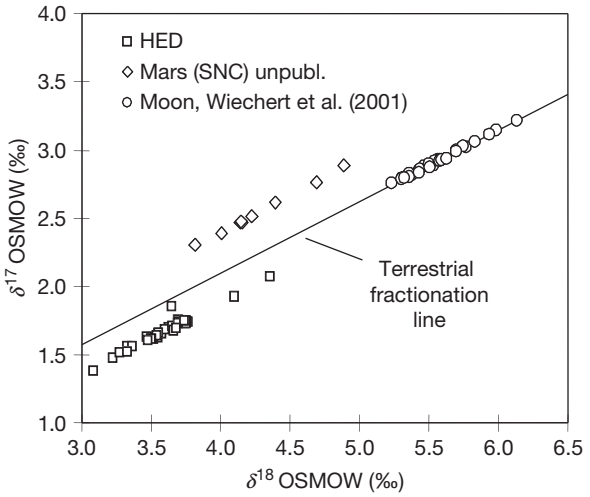


Figure 3 The oxygen isotopic composition of the components in chondrites, in particular CAIs, is highly heterogeneous for reasons that are unclear. The net result of this variability is that different planets possess distinct oxygen isotopic compositions that define an individual mass fractionation line as shown here for eucrites, howardites, and diogenites, which come from Vesta and martian (SNC) meteorites, thought to come from Mars. The Moon is thought to have formed from the debris produced in a giant impact between the proto-Earth when 90% formed and an impacting Mars-sized planet sometimes named 'Theia.' The fact that the data for lunar samples are collinear with the terrestrial fractionation line could mean that the Moon formed from the Earth, or the planet from which it was created was formed at the same heliocentric distance, or it could mean that the silicate reservoirs of the two planets homogenized during the impact process, for example, by mixing in a vapor cloud from which lunar material condensed. Reproduced from Halliday AN (2003) The origin and earliest history of the Earth. In: Davis AM (ed.) *Meteorites, Comets and Planets*, Holland HD and Turekian KK (eds.) *Treatise of Geochemistry*, vol. 1, pp. 509–557. Oxford: Elsevier-Pergamon.

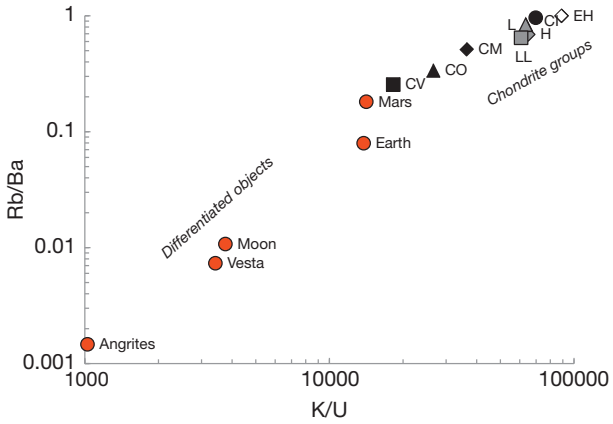


Figure 4 Comparison between the K/U and Rb/Bar ratios of the Earth and other differentiated objects compared with chondrites. The alkali elements K and Rb are both relatively volatile compared with U and Ba that are refractory. Therefore, these trace element ratios provide an indication of the degree of volatile element depletion in inner solar system-differentiated planets relative to chondrites, which are relatively primitive. It can be seen that the differentiated objects are more depleted in moderately volatile elements than are chondrites. Reproduced from Halliday AN (2014) The origin and earliest history of the Earth. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, 2nd edn., vol. 2, pp. 149–211. Oxford: Elsevier.

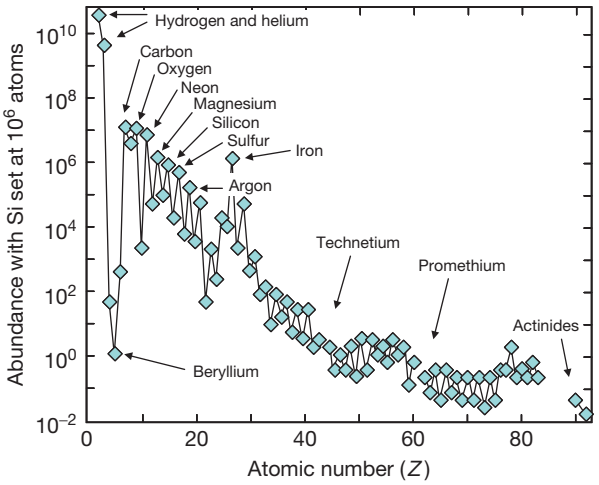


Figure 5 The abundances of elements in our Sun and the solar system are estimated from the spectroscopic determination of the composition of the Sun and the laboratory analysis of primitive meteorites called carbonaceous chondrites – thought to represent unprocessed dust and other solid debris from the circumsolar disk. To compare the abundances of different elements, it is customary to scale the elements relative to one million atoms of silicon. The pattern provides powerful clues to how the various elements were created. See text for details. Based on a figure in Broecker WS (1998) *How to Build a Habitable Planet*. Columbia University, New York: Eldigio Press with kind permission.

abundance at Fe, which is about 1000 times more abundant than its neighbors, reflecting the maximum binding energy per nucleon at ^{56}Fe. There is a big dip in the abundance of Li, Be, and B because these elements are not stable in stellar interiors. There is also a saw-toothed variability that is superimposed on the overall curve, reflecting the relatively high stability of even numbered nuclides compared with odd numbered nuclides. Finally, all of the elements in the periodic table are present in the solar system except those with no long-lived or stable nuclides. Tc, promethium (Pm), and the transuranic elements fall into this category.

The four most abundant elements in the terrestrial planets are O, Mg, Si, and Fe. Broadly speaking, the density of the Earth and the size of its core provide a first-order indication that the ratio of Fe to Mg is at least similar to the values found in chondrites. In fact, it is slightly higher, as discussed later in the text, but to a first approximation, the Earth has chondritic proportions of elements that are refractory or only slightly volatile. Since carbonaceous chondrites appear to represent undifferentiated protoplanetary material, it seems likely that the bulk Earth contains the same relative proportions of all refractory elements as carbonaceous chondrites. This is the basis of the *chondritic reference model*. In detail, there are significant differences relative to chondrites, particularly with respect to the volatile elements. Furthermore, core formation has redistributed the Earth's elements internally. During and after growth of the Earth, *siderophile* (literally 'metal-loving') elements were partitioned into the metallic core, leaving *lithophile* (literally 'rock-loving') elements behind in the silicate mantle.

The composition of the *bulk silicate Earth* (BSE), otherwise known as the *primitive mantle* (PM), can be deduced in more detail by determining the compositions of a wide range of basalts and/or peridotite xenoliths and using these to identify specific features that are systematic. Determining BSE in this way requires that one makes an assumption of a homogeneous mantle. This assumption is still hotly debated by some, and for more detail, one can refer to companion volumes 2 (Mineral Physics) and 7 (Mantle Dynamics). However, two key observations lend confidence that the chemistry of the mantle is broadly speaking homogeneous, in striking contrast to Mars, for example (Lee and Halliday, 1997). First, no isotopic heterogeneity older than about 2 billion years has survived in the oceanic mantle, pointing to widespread efficient mixing processes over geologic time. Second, the ratios of the refractory lithophile elements Nb/U are relatively uniform but nonchondritic and fractionated by crust production (Hofmann et al., 1986). There is no reason why this ongoing fractionating process should have left the mantle so homogeneous unless the mantle is also relatively well mixed.

Assuming these features are representative of the total mantle, one can combine estimates of mantle composition with those of the average composition of the continental crust to define the composition of the PM (=BSE). For example, mantle peridotites are fractionated residues of partial melting but ratios of moderately or highly compatible elements such as Mg/Fe or concentrations such as of Ni or Mg can be compared to a wide range of samples and a representative estimate made for the composition of the BSE.

A different approach is adopted for the highly incompatible elements, which are lost from peridotite during partial melting. Basaltic magmas are the products of partial melting of broad regions of the peridotitic mantle. They fractionate compatible elements one from another but have lesser effects on highly incompatible elements such as the large ion lithophile elements. The latter are essentially completely extracted into the basaltic melt, yielding constant ratios one to another unless the degree of melting is very small. Element ratios that are not significantly fractionated by partial melting, as evidenced by little or no variation in ratio with change in concentration, are particularly useful for determining the BSE depletion in moderately volatile elements. The volatile element depletion can be determined from their ratio relative to a refractory element of similar incompatibility if both elements are lithophile and hence not partitioned into the core. Relevant examples include Rb/Ba (Hofmann and White, 1983), K/U (Jochum et al., 1983), and In/Y (Yi et al., 1995, 2000). A similar approach can be adopted for siderophile elements. Refractory siderophile/refractory lithophile element ratios for pairs of elements of similar incompatibility provide a clear indication of depletion caused by core formation, for example, Mo/Pr and W/Ba (Newsom et al., 1986). Finally, one can in the same way deduce the PM concentrations of some elements that are depleted because of both core formation and volatile loss, for example, ^{204}Pb/Ce (Hofmann et al., 1986), Sn/Sm (Jochum et al., 1993; Yi et al., 2000), and Cd/Dy (Yi et al., 2000). As with peridotite data, the budgets for the continental crust need to be added back in with those determined for the composition of the (depleted) mantle, to derive a value for the PM (which is the same as the BSE).

The conclusion reached from determining relative proportions of less incompatible lithophile elements in peridotites such as heavy rare earth elements (REEs), Ca, Sc, and Ti (e.g., Lee et al., 1996) is that, although the bulk Earth does not have an exactly CI composition (Drake and Righter, 2002), refractory lithophile elements are present in approximately the same relative proportions in the BSE as in the carbonaceous chondrites (Allègre, et al., 1995b; McDonough, 2003; McDonough and Sun, 1995). This view has been challenged because of the Earth's nonchondritic Fe/Mg and apparent Nd isotopic composition (discussed later in the text).

The moderately volatile elements such as potassium (K) are those predicted to have condensed 50% of their mass from a gas of solar composition at temperatures of 700–1100 K (Lodders, 2003). They are depleted in most chondrite groups when compared to solar or CI compositions (**Figures 4** and **6**). There has been much discussion in the literature of the cause of this, usually related to imperfect mixing of CAI, chondrule, and matrix fractions (e.g., Zanda et al., 2006). Moderately volatile elements are strongly depleted in the Earth and other differentiated objects, more so even than in chondrites (**Figures 4** and **6**). By plotting volatile siderophile/refractory lithophile element ratios (e.g., S/Sr) in the BSE and chondrites versus ratios of lithophile moderately volatile to refractory lithophile elements (e.g., Rb/Sr), one can estimate the likely partitioning of volatile elements into the core (Halliday and Porcelli, 2001) (**Figure 6**).

The very refractory lithophile elements are enriched relative to chondrites by a factor of ~3 (McDonough and Sun, 1995), more than can be explained by the separation of the core. A plot of concentrations of elements in the BSE divided by those in CI chondrites, normalized to $[Mg]_{Earth}/[Mg]_{CI} = 1.0$ in order to correct for the highly volatile contents of CI chondrites, is shown in **Figure 7**. These elemental ratios are plotted as a function of the temperature by which 50% of the element of interest would have condensed during the cooling of a gas of solar system composition (Lodders, 2003). Note that condensation may not be the mechanism of volatile loss.

If one compares the BSE composition, determined as described earlier in the text, with that of the Sun or CI chondrites (**Figures 4** and **6–8**), several features stand out:

1. There is an overall trend of decreasing chondrite-normalized abundance of lithophile elements as a function of decreasing half-mass condensation temperature (**Figure 7**). Losses of volatiles were important.
2. Some elements, for example, the halogens, are depleted (**Figure 7**) for reasons that are less clear. Even the alkalies do not display a clear relationship between depletion and volatility. Therefore, volatile losses may have involved additional controls beyond nebular condensation temperature.
3. The siderophile elements are depleted, even if they are refractory, presumably because they are partitioned into the Earth's core. The most depleted, hence presumably the most siderophile, are the platinum group elements (PGEs), rhenium (Re), sulfur (S), selenium (Se), and tellurium (Te) (**Figure 7**). They are present in roughly chondritic-relative proportions, a striking feature that relates to the history of accretion and core formation, discussed later in the text.
4. The Si/Mg ratio of the BSE is lower than is found in chondrites (**Figure 8**). A variety of reasons have been proposed

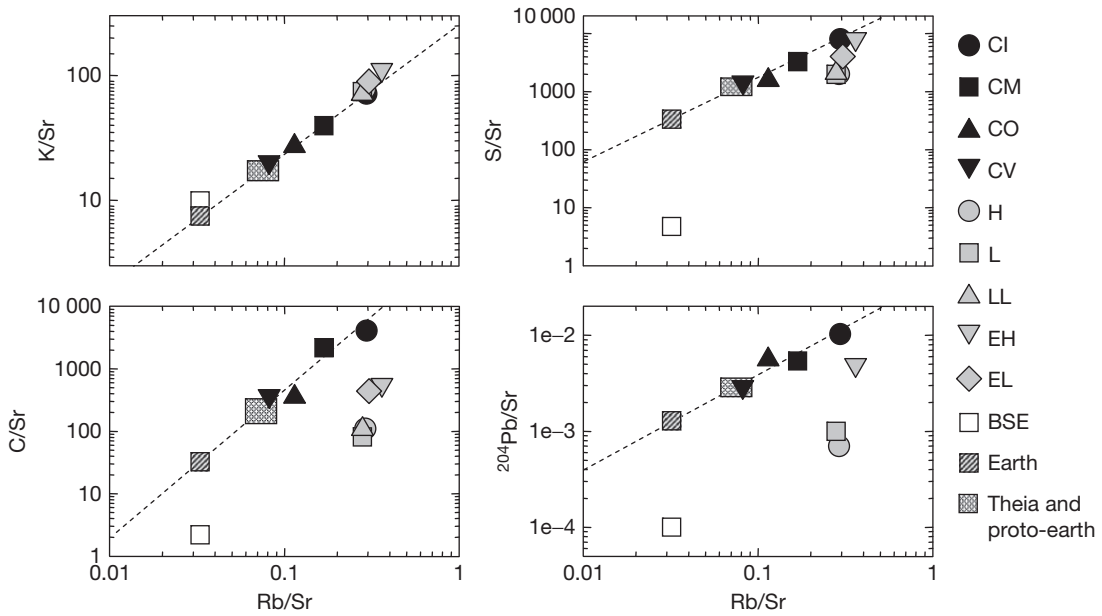


Figure 6 Volatile/refractory element ratio–ratio plots for chondrites and the silicate Earth. The correlations for carbonaceous chondrites can be used to define the composition of the Earth, the Rb/Sr ratio of which is well known because the Sr isotopic composition of the bulk silicate Earth (BSE) represents the time-integrated Rb/Sr. Data from Newsom HE (1995) Composition of the solar system, planets, meteorites, and major terrestrial reservoirs. In: *Global Earth Physics: A Handbook of Physical Constants*. Washington, DC: American Geophysical Union, AGU Reference Shelf, No. 1.

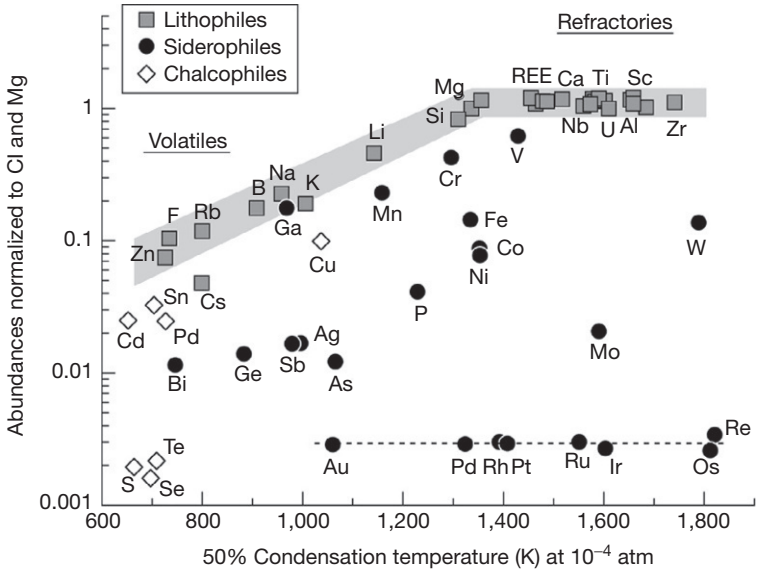


Figure 7 Elemental abundances in the silicate Earth (McDonough and Sun, 1995) ratioed to those in CI carbonaceous chondrites and normalized to $[Mg]_{Earth}/[Mg]_{CI} = 1.0$ are plotted on the ordinate. The abundances are plotted against the temperature by which 50% of the element of interest would have condensed from a gas of solar composition at a total pressure of 10^{-4} bar (Lodders, 2003). The silicate Earth is depleted in volatile elements (which condense at low temperatures) due to a combination of incomplete accretion and late volatile loss. Depletions of the silicate Earth in siderophile and some chalcophile elements are due to sequestration in the core.

to account for this, including partitioning of Si into the core (Gessmann et al., 2001; Takafuji et al., 2005; Wade and Wood, 2005), losses of Si to space (Ringwood, 1989a,b), and the idea that the Earth did not actually accrete from material that was the same as chondrites (Drake and Righter, 2002; Hewins and Herzberg, 1996).

5. The Mg/Fe ratio of the total Earth also appears to be slightly lower than is found in chondrites (Palme, 2000). The reason for this is also unclear. It could simply be that the Earth formed from nonchondritic material but it is also possible that this reflects some level of erosion of the outer part of the Earth during accretion (O'Neill and Palme, 2008), as

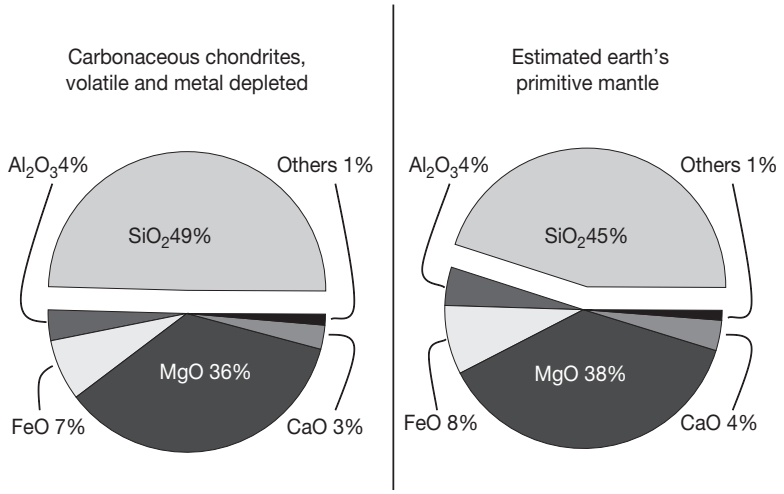


Figure 8 Pie chart showing the composition of the BSE compared with CI chondrites after correction for differences in volatiles and subtraction of a 'core' component. Note the depletion in Si that may be caused by partitioning into the core or a nonchondritic Earth.

has been proposed for Mercury, with its high density, hence proportionally bigger core (Benz et al., 1987).

The reasons for these patterns in composition of the Earth remain the focus of considerable debate. They provide evidence that the growth of the Earth was far from simple and may have changed with time, a view reinforced by isotopic data. However, a part of the composition probably relates to the Earth's location within the circumstellar disk from which the raw materials were obtained. This is therefore discussed first before moving on to other aspects related to the development of the Earth itself.

9.02.6 The Circumstellar Disk and the Composition of the Earth

About 200 years ago, Kant and Laplace considered that the most likely explanation for the formation of the solar system was planetary accretion from a swirling disk resulting from the angular momentum of the cloud of gas and dust that collapsed to form the Sun. A number of the previously mentioned compositional features of the Earth are thought to relate to early processes associated with such a circumstellar disk.

The general explanation for the depletion of the Earth in highly volatile elements such as H and He is that most of the growth of terrestrial planets, unlike the gas giants, post-dated the loss of nebular gases from the disk. However, this is far from certain. For example, a recent model for the origin of the Earth's water is that it was formed from nebular hydrogen trapped by the Earth as it grew (Ikoma and Genda, 2006). It is not possible to trap significant gas in this way until the Earth reaches >30% of its current mass, which given the short timescales for nebula dispersal is hard to reconcile with protracted accretion (see later text). Ikoma and Genda pointed out that the nebula can be much less dense than the nominal value, that is, captured gas is only logarithmically sensitive to nebula density, although the laws governing nebula dispersal and declining gas density are not well established.

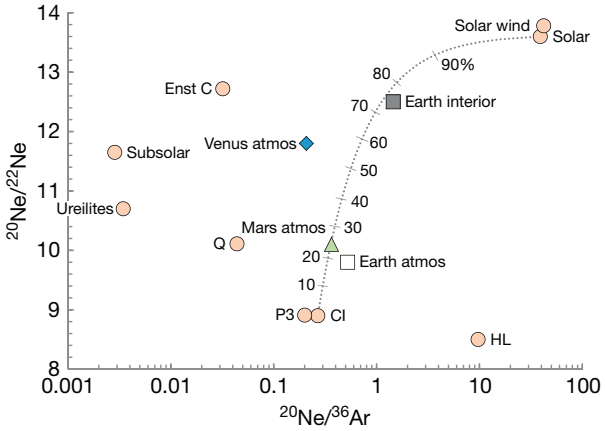


Figure 9 Neon isotopic compositions for the atmospheres of terrestrial planets are no different from the range of isotopic compositions in the solar system. There is no need to advocate fractionation caused by atmospheric loss. Indeed, the neon isotopic compositions of the terrestrial planets correlate negatively in a non-mass-dependent fashion with $^{20}Ne/^{36}Ar$. The variation is better explained in terms of admixing differing amounts of solar composition material with noble gases derived from CI chondrite-like material. On this basis, the mantle has a proportion of solar Ne of 75%, whereas in the atmosphere, this proportion is 20%. Reproduced from Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.

In support of the Ikoma and Genda model, there is clear isotopic evidence that some solar noble gases were trapped in the Earth (Porcelli et al., 2001) and Mars (Marty and Marti, 2002), possibly as a vestige of the nebula still present at the time of the Earth's accretion (Hayashi et al., 1979). Whether some is stored in the core is unclear (Porcelli and Halliday, 2001). The amount of solar Ne in the mantle appears to be ~75% (**Figure 9**) but is small compared with the far larger (×100) reservoir of atmospheric Ne, more than 80% of which is chondritic (Halliday, 2013). The relative size of the

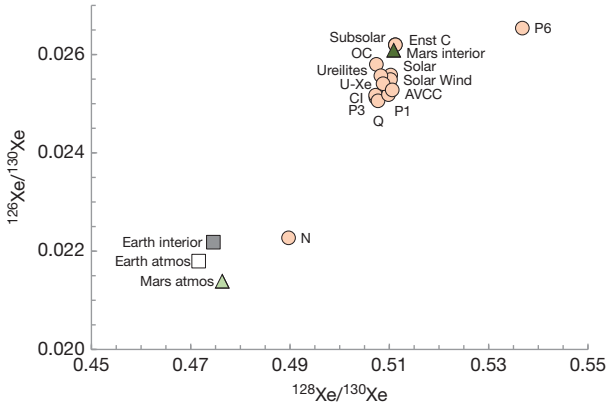


Figure 10 Xenon is isotopically fractionated in the Earth's atmosphere and mantle relative to the estimated composition of the material from which it formed. A similar heavy isotopic composition is found in the martian atmosphere. Reproduced from Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.

atmosphere versus interior budgets of noble gases is deduced from the decay of ^{40}K–^{40}Ar combined with the mantle $^{40}Ar/^{36}Ar$ and assuming that the mantle as sampled by ocean ridge basalts, ocean island basalts, and well gases is representative (Halliday, 2013). On this basis, <3% of the Earth's ^{36}Ar resides in the mantle. There is Xe isotopic evidence discussed later in the text that the majority (>99%) of the Earth's noble gases were lost at a late stage (Porcelli and Pepin, 2000) (**Figure 10**) possibly during the Moon-forming giant impact. This is consistent with the difference in noble gas budgets between the Earth and Venus (Halliday, 2013). However, the isotopic composition of the nonfissionogenic and nonradiogenic Xe in the Earth looks like neither fractionated Solar nor fractionated chondritic-relative abundances. Its origin is an enigma (Pepin, 2000). The dynamics and timescales for accretion are dependent on the presence or absence of nebular gas as discussed in the succeeding text.

As already pointed out, the moderately volatile elements are also depleted in the Earth (**Figures 4, 6**, and **7**) (Cassen, 1996; Gast, 1960; Wasserburg et al., 1964). The traditional explanation for this is that the inner 'terrestrial' planets accreted where it was hotter, within the so-called ice line (Cassen, 1996; Humayun and Cassen, 2000). It long was assumed for several reasons that the solar nebula in the terrestrial planet-forming region started as a very hot, well-mixed gas from which all of the solid and liquid Earth materials condensed. The geochemistry literature contains many references to this hot nebula and to major T-Tauri heating events that may have further depleted the inner solar system in moderately volatile elements (e.g., Lugmair and Galer, 1992). Some nebula models predict early temperatures that were sufficiently high to prevent condensation of moderately volatile elements (Humayun and Cassen, 2000) that somehow were lost subsequently.

Understanding how the volatile depletion of the inner solar system occurred can be greatly aided by comparisons with modern circumstellar disks. These were first detected using ground-based interferometry. Circumstellar protoplanetary disks (termed *proplyds*) now are plainly visible around young stars in the Orion Nebula thanks to the Hubble Space Telescope (McCaughrean and O'Dell, 1996). With such refined observations combined with theoretical considerations, it is possible to trace the likely early stages of development of our own solar system (Ciesla and Charnley, 2006).

A solar T-Tauri stage during the development of the circumstellar disk could have had a profound effect on volatile depletions. T-Tauri stars are premain sequence stars. They are a few times 10^5 years to a few times 10^6 years in age and have many of the characteristics of our Sun but are much brighter. Many have disks. It has long been argued that the T-Tauri effect causes an early phase of heating of the inner portions of the disk. The T-Tauri stage may last a few million years and is often linked to the Earth's depletion in moderately volatile elements. Because it heats the disk surface however, it is unsure whether it would have a great effect on the composition of the gas and dust in the accretionary midplane of the disk where planetesimal accretion is dominant. Heating of inner solar system material in the midplane of the disk will, however, occur from compressional effects. The thermal response can be calculated for material in the disk being swept into an increasingly dense region during migration toward the Sun during the early stages of disk development. Boss (1990) included compressional heating and grain opacity in his modeling and showed that temperatures in excess of 1500 K could be expected in the terrestrial planet-forming region. The main heating takes place at the midplane, because that is where most of the mass is concentrated. The surface of the disk is much cooler. This process would certainly be very early but how it would have evolved with time and tie into the timescales for disk heating deduced from cosmochemistry is unclear.

Chondrules and CAIs provide us with the most important information about the processes of heating and melting in the disk that occurred within the first few million years of the solar system. The exact nature of these events has been hard to establish but chondrules are most likely produced by flash melting of the dusty disk (Kita et al., 2006). What role they played in the formation of planetesimals has been unclear, but some models predict that they might have accumulated preferentially in certain stagnant regions of the disk, facilitating planet formation (Cuzzi et al., 2001). Hewins and Herzberg (1996) had proposed that this is reflected in the Earth's nonchondritic Si/Mg (**Figure 8**).

9.02.7 Dynamics of Planet Formation

In broad terms, the rates of accretion of planets from disks are affected by the amounts of mass in the disks themselves. If there is considerable nebular gas present at the time of accretion, the rates are faster (e.g., Hayashi et al., 1985). If gas is present, then it is important to know what Jupiter is doing at the time of terrestrial planet accretion. At present, this is underconstrained and model-dependent. The absence of nebular gas is also calculated to lead to eccentric orbits (Agnor and Ward, 2002). Recently, however, it has been shown that accretion in the presence of swarms of planetesimals also decreases eccentricity (O'Brien et al., 2006).

The most widely accepted model of terrestrial planet formation is the planetesimal theory (Chambers, 2004). In the simplest terms, accretion of terrestrial planets is envisaged as taking place in four stages:

1. Settling of circumstellar dust to the midplane of the disk
2. Growth of planetesimals up to ~1 km in size
3. Runaway growth of planetary embryos up to $\sim 10^3$ km in size
4. Stochastic growth of larger objects through late-stage collisions

Stage 1 takes place over timescales of thousands of years if there is no turbulence but much longer otherwise. It provides a relatively dense plane of material from which the planets grow. The second stage is poorly understood but is necessary in order to build objects that are of sufficient mass for gravity to play a major role. Planetesimals would need to be about a kilometer in size in order for the gravitationally driven stage 3 to start. Although we do not know how stage 2 happens, somehow, it must be possible. Fluffy aggregates of dust have been made in the lab but these are typically less than a centimeter in size (Blum, 2000). Larger objects are more problematic. One obvious suggestion is that some kind of glue was involved. Volatiles would not have condensed in the inner solar system, however. Not only were the pressures too low, but also the temperatures were probably high because of heating as material was swept into the Sun (Boss, 1990). Organic material or molten droplets such as chondrules (Wiechert and Halliday, 2007) may have played an important role in cementing material together. An alternative to glues that has long been considered (Ward, 2000) is that within a disk of dust and gas, there are local separation and clumping of material as it is being swept around. This leads to gravitational instabilities whereby an entire section of the disk has relatively high gravity and accumulates into a zone of concentrated mass (Goldreich and Ward, 1973; Weidenschilling, 2006). This is a similar to some models envisaged by some for Jupiter formation (Boss, 1997).

Once stage 2 has occurred, runaway growth builds the 1 km-sized objects into 1000 km-sized objects. This mechanism exploits the facts that (1) larger objects exert stronger gravitational forces, (2) the velocity dispersion is small so the cross section for collision can be much larger than the physical cross section for the largest bodies in the swarm, (3) collisions resulting in growth are favored if the material is not on an inclined orbit, and (4) larger objects tend not to take inclined orbits. The net result is that the bigger the object, the larger it becomes until all of the material available within a given feeding zone or heliocentric distance is incorporated into planetary embryos. This is thought to take place within a few hundred thousand years (Chambers, 2004; Kortenkamp and Wetherill, 2000; Kortenkamp et al., 2000; Lissauer, 1987, 1993). The ultimate size depends on the amount of material available. Using models for the density of the solar nebula, it is possible that a body of the size of Mars ($\sim 0.1 M_E$) could originate in this fashion at its current heliocentric distance. However, in the vicinity of the Earth, the maximum size of object would be Moon-sized ($\sim 0.01 M_E$) or possibly even smaller.

Building objects that are as large as the Earth is thought to require a more protracted history of collisions between the 1000 km-sized planetary embryos (Wetherill, 1986). This is a stochastic process that is hard to predict in a precise manner. The Russian theoretician Safronov (1954) proposed that, in the absence of a nebula, the growth of the terrestrial planets would be dominated by such a history of planetary collisions. Wetherill (1980, 1986) took this model and ran Monte Carlo simulations of terrestrial planetary growth. He showed that indeed, some runs generated planets of the right size and distribution to be matches for Mercury, Venus, the Earth, and Mars. He monitored the timescales involved in these 'successful' runs and found that most of the mass was accreted in the first 10 Ma but that significant accretion continued for much longer. Wetherill also tracked the provenance of material that built the terrestrial planets and showed that, in contrast to runaway growth, the feeding zone concept was flawed. The planetesimals and planetary embryos that built the Earth came from distances that extended over more than 2 AU. More recent calculations of solar system formation have yielded similar results (Canup and Agnor, 2000).

The latest 'Grand Tack' models for the solar system's accretion propose that *all* of the terrestrial planets and asteroids forming inboard of 1 AU in a relatively confined zone (Hansen, 2009). This helps to explain the small size of Mars. The new standard model of the inner solar system is that planetary embryos such as iron meteorite parent bodies accreted close into the Sun but then migrated outward subsequently (Bottke et al. (2006). The 1 AU edge of terrestrial planet accretion was set by an inward-migrating Jupiter that cleared or accreted everything in its path. However, Jupiter was then slowed and subsequently pulled back by the growth of Saturn. The terrestrial planets and asteroid belt spread out across the inner disk as these two giant planets migrated back outward (Hansen, 2009; Walsh et al., 2011).

Planetary collisions of the type that grew the terrestrial planets would have been catastrophic. The energy released would have been sufficient to raise the temperature of the Earth by thousands of degrees. The most widely held theory for the formation of the Moon is that there was such a catastrophic collision between a Mars-sized planet and the proto-Earth when it was approximately 90% of its current mass. The putative impactor planet sometimes named 'Theia' (the mother of 'Selene' who was the goddess of the Moon) struck the proto-Earth with a glancing blow generating the angular momentum of the Earth–Moon system.

These dynamic models can be tested with geochemistry and petrology, which provide 6 principal kinds of information relevant to the earliest history of the Earth:

1. Isotopic heterogeneity in the solar system can be used to trace the sources of the components that built the Earth and Moon.
2. Timescales for accretion, volatile loss, and core formation can be determined from short- and long-lived decay systems (Figure 1).
3. Conditions of core formation can be determined from metal–silicate partitioning behavior of indicative 'siderophile' elements.
4. Time-integrated parent/daughter ratios can be determined from isotopic compositions of the daughter elements yielding insights into the (paleo)chemistry of precursor materials.

5. Mass-dependent stable isotopic fractionations provide constraints on the prevalent processes and conditions.
6. Geochemistry and petrology provide our principal constraints on the state of the Earth at the end of accretion.

These represent a huge area of science and a continually evolving landscape as new and improved techniques are brought to bear on meteorites and lunar samples. The key aspects as they relate to the early Earth are as follows.

9.02.8 The Age of the Earth

Earth's antiquity is a problem that in many respects was solved when Clair Patterson published the results of Pb isotopic analyses of meteorites that lay on an isochron passing near the composition of the silicate Earth (**Figure 11**) and indicating an age of 4.5–4.6 Ga (Patterson, 1956). Following this work, many refinements to the Earth's age were suggested based on better estimates of the average Pb isotopic composition of the BSE (Galer and Goldstein, 1996) but the essential result remains unchanged. These further studies highlight the fact that there are a variety of inherent complications to take into account in determining an isotopic age or rate. These are very relevant to the issue of how well we know the exact timescales for the Earth's formation (Halliday, 2014).

In most isotopic dating, one has to make allowance for the initial daughter isotope that was present in the mineral, rock, or reservoir when it formed. This can be corrected for adequately or ignored in certain minerals with very high parent element/daughter element ratios. For example, the most accurate ages of ancient rocks in the continental crust are obtained by U–Pb dating of the common zirconium silicate mineral zircon. This mineral has high U/Pb ratio when it forms and is resistant to the effects of natural resetting via diffusion. Lead does not fit readily into the lattice when zircon grows so any initial Pb is minor and corrections can be made with sufficient confidence that the ages are extremely accurate. The oldest portion of a terrestrial zircon grain thus far found yields an age of 4.40 Ga (Wilde et al., 2001).

There are no rocks surviving from the first 500 Ma of the Earth's history and no zircons from the first 100 Ma. The age of the Earth and the solar system is determined using a different approach. Most minerals and rocks carry a significant proportion of initial Pb relative to that formed by radioactive decay. It is necessary to monitor the initial abundance of the daughter isotope precisely by measuring the parent and daughter isotopes relative to a nonradiogenic isotope of the same daughter element. This is commonly done using isochrons that allow one to determine an age without knowing the initial composition a priori. There are two isotopes of U that decay to two isotopes of Pb. By comparing the Pb isotopic ratios in reservoirs that all formed at the same time, it is even unnecessary to know the parent/daughter ratios, since the ratio $^{238}U/^{235}U$ is constant in nature (137.88). It was using these techniques that Clair Patterson determined that the age of iron and silicate meteorites, and the Earth, was between 4.5 and 4.6 billion years. This defines the time at which the major variations in U/Pb between U-depleted iron meteorites and (variably) Pb-depleted silicate reservoirs, such as the Earth's PM, were established by volatile loss and core formation (**Figure 12**).

Although Pb–Pb dating was at one time used for dating a wide variety of terrestrial samples, this method is all but obsolete except for studying extraterrestrial samples. Most chondrites contain CAIs, which, as previously explained, are enriched in elements expected to condense at very high temperatures from a hot nebular gas, such as U. These are the oldest objects yet identified that formed in the solar system. CAIs from the Efremovka chondrite have been dated by $^{235/238}U–^{207/206}Pb$ at 4.5672 ± 0.0006 Ga (Amelin et al., 2002), and although minor revisions have been proposed, this is still the best estimate of the age of the solar system (**Figure 2**) that defines a more precise slope to the meteorite isochron (called the 'Geochron') first established by Patterson (**Figure 11**).

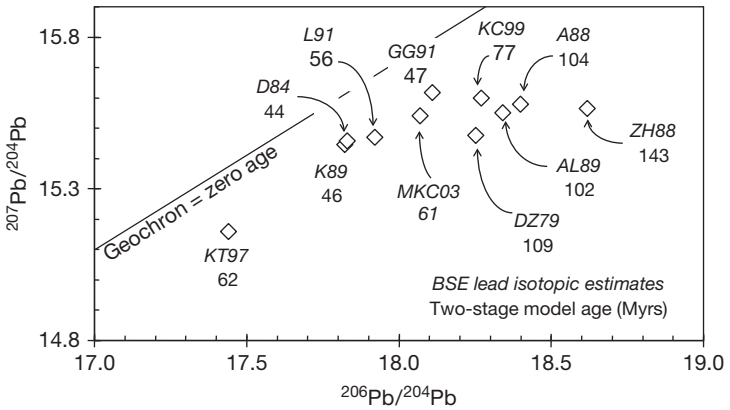


Figure 11 Estimates of the lead isotopic composition of the BSE plotted relative to the Geochron defined as the slope corresponding to the start of the solar system. All estimates plot to the right of this line, which, if correct, indicate protracted accretion and/or core formation. The times indicated in Ma are the two-stage model ages of core formation assuming the same values for bulk Earth parameters given in Halliday (2004) and Wood and Halliday (2005). Data from Doe and Zartman (1979), Davies (1984), Zartman and Haines (1988), Allègre et al. (1988), Allègre and Lewin (1989), Kwon et al. (1989), Liew et al. (1991), Galer and Goldstein (1991), Kramers and Tolstikhin (1997), Kamber and Collerson (1999), and Murphy et al. (2003). These Pb isotopic estimates are all significantly longer than the $^{182}Hf–^{182}W$ estimate of 30 Ma (Kleine et al., 2002; Yin et al., 2002).

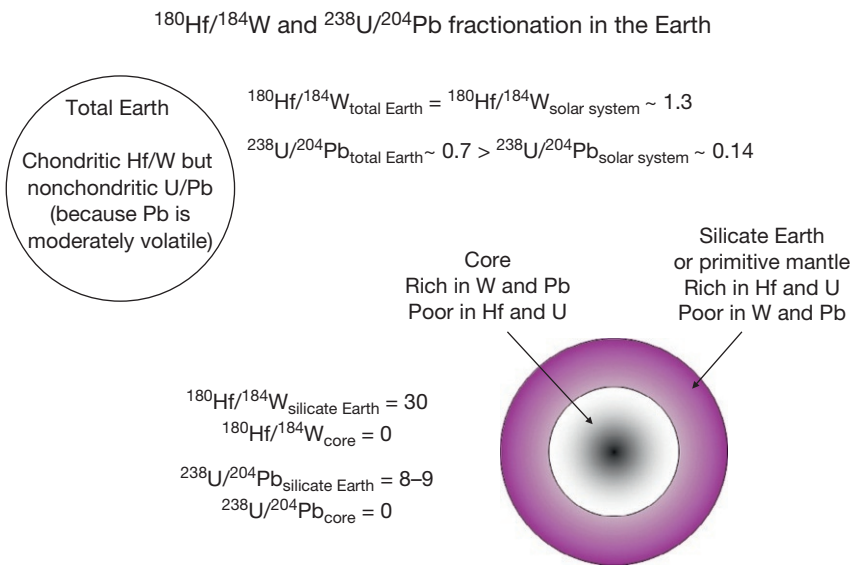

$^{180}Hf/^{184}W$ and $^{238}U/^{204}Pb$ fractionation in the Earth
Total Earth
Chondritic Hf/W but nonchondritic U/Pb (because Pb is moderately volatile)
$^{180}Hf/^{184}W_{total\ Earth} = {}^{180}Hf/^{184}W_{solar\ system} \sim 1.3$
$^{238}U/^{204}Pb_{total\ Earth} \sim 0.7 > {}^{238}U/^{204}Pb_{solar\ system} \sim 0.14$
Core
Rich in W and Pb
Poor in Hf and U
Silicate Earth
or primitive mantle
Rich in Hf and U
Poor in W and Pb
$^{180}Hf/^{184}W_{silicate\ Earth} = 30$
$^{180}Hf/^{184}W_{core} = 0$
$^{238}U/^{204}Pb_{silicate\ Earth} = 8–9$
$^{238}U/^{204}Pb_{core} = 0$

Figure 12 Hafnium–tungsten chronometry provides insights into the rates and mechanisms of formation of the solar system, whereas U–Pb chronometry provides us with an absolute age of the solar system. In both cases, the radioactive parent/radiogenic daughter element ratio is fractionated by core formation, an early planetary process. It is this fractionation that is being dated. The Hf/W ratio of the total Earth is chondritic (average solar system) because Hf and W are both refractory elements. The U/Pb ratio of the Earth is enhanced relative to average solar system because approximately >80% of Pb was lost by volatilization or incomplete condensation mainly at an early stage of the development of the circumstellar disk. The fractionation within the Earth for Hf/W and U/Pb is similar. In both cases, the parent (Hf or U) prefers to reside in the silicate portion of the Earth. In both cases, the daughter (W or Pb) prefers to reside in the core.

Since Patterson published his work, various researchers have tried to estimate the average Pb isotopic composition of the bulk silicate portion of the Earth. This provides information on the timing of U/Pb fractionation caused by terrestrial core formation (Figure 12). The results of these various estimates are shown in Figure 11. It can be seen that none plot on the Geochron and this provide evidences that accretion of the Earth, or core formation, or both, was either late or protracted. These results need to be interpreted with caution for two reasons. Firstly, the average Pb isotopic composition of the bulk silicate portion of the Earth is very hard to estimate because the long-lived decay of U has resulted in considerable isotopic heterogeneity. All of the estimates could in principle be wrong. Secondly, converting these results into timescales is strongly model-dependent involving assumptions about how the Earth formed. The results of one simple model are shown in Figure 11. In this model, it is assumed that no U/Pb fractionation took place until a single point in time. So, for example, this could describe an Earth that formed its entire core in a single event (Figure 13(a)). The ages defined by the Pb isotopic compositions correspond to tens of millions of years after the start of the solar system (Figure 11). Although such data were at one time used to 'date' core formation, it is now recognized that the processes of accretion and core formation are more complex than can be described by such an approach. The data do, however, provide evidence of a protracted history of accretion and core formation. To sort out that history, it is necessary to use short-lived nuclides, dynamic simulations of planet formation, and petrologic constraints on likely core formation scenarios.

9.02.9 Short-Lived Nuclides and Early Processes

Since Patterson's study, it has become clear that one can resolve relatively small time differences between early solar system objects using more precise and different techniques (Figure 1). These provide us with fascinating information on accretion rates and permit the determination of which dynamic models of solar system formation are most probably correct. Progress in this field has been huge in the past 10 years. It is now becoming clear that different isotopic systems applied to the Earth most probably 'date' different processes. Volatile loss, metal segregation, sulfide segregation, mantle melting, atmosphere formation, and continental growth all affect particular elements in distinctive ways. As such, the isotopic geochemistry of an expanding range of elements is facilitating the first comprehensive view of the geochemical evolution of the early Earth.

The short-lived nuclides listed in Table 1 provide a set of powerful tools for unraveling a precise chronology of the early solar system. The advantage of these isotopic systems is that the changes in daughter isotope can only take place over a restricted early time window and there is no correction for the effects of decay over the past 4.5 billion years. They were present in the early solar system because of production in other stars, many of them massive, shortly before (within a few half-lives of) the formation of the Sun (Figure 1). Being extinct, the initial abundance can only be inferred. This is achieved by using differences in the atomic abundance of the daughter isotope in early objects (meteorites) of known absolute age.

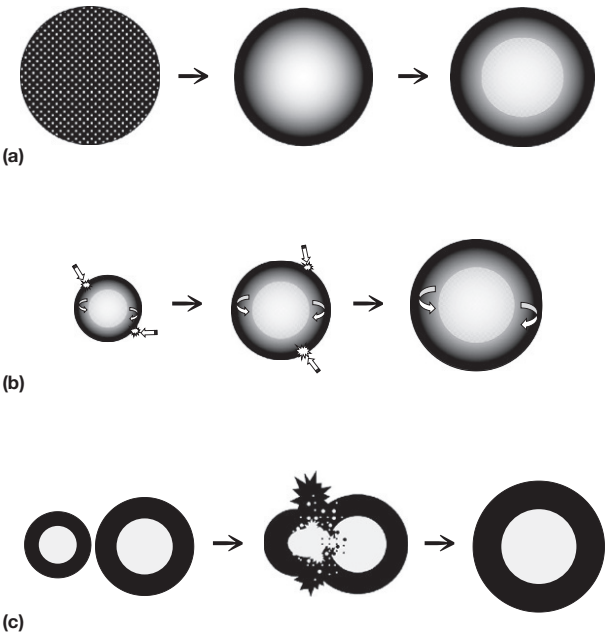


Figure 13 Schematic representation of core formation scenarios and how they translate into isotopic models and timescales. (a) *Instantaneous core formation*. In this model, an age is calculated as the point in time that a fully formed planet segregated into core and silicate reservoirs. (b) *Continuous core formation*. The planet accretes from material with a chondritic Hf/W ratio and W isotopic composition. Newly accreted material is continuously mixed with W in the outer silicate portion prior to the segregation of new iron and siderophile elements (including W) into the core. Most models assume an exponentially decreasing rate of the overall growth for the planet and its core. (c) *Core–core mixing*. Terrestrial planets appear to have formed from material that was already differentiated into silicate and metal. How much equilibration between accreted metal and the planet's silicate reservoir occurs (an explicit requirement in calculating an age with model (b)) is not well understood at present. The net result if calculated with continuous core formation and equilibration as in model (b) is that the rate of growth of the planet appears faster than it really was. Reproduced from Halliday AN (2006) The origin of the Earth – What's new? *Elements* 2: 205–210.

To utilize any of the nuclides of Table 1 for deducing early solar system timescales, it is necessary to measure precisely the present-day isotopic composition of the daughter element. Therefore, the development of these chronometers is strongly linked to breakthroughs in mass spectrometry. In the simplest system, the present-day isotopic composition of the daughter element is a function of the initial abundance of the parent and daughter nuclide at the start of the solar system, the parent/daughter element ratio, and how much time elapsed. In effect, the isotopic composition of the daughter represents a time-integrated parent/daughter elemental ratio. If an object forms without a change in parent/daughter ratio, its time of formation will be undetectable with isotopic chronometry. It is the large change in parent/daughter ratio that accompanies certain processes associated with planetary accretion that renders some such isotopic systems useful. In fact, it is this change that is being dated. If there is no change that accompanies accretion, there is no way to distinguish when an object formed relative to the start of the solar system.

The ^{182}Hf–^{182}W chronometer (Table 1) has been particularly useful in determining the timescales over which the planets formed. The principles are as follows: Excess ^{182}W comes from the radioactive decay of ^{182}Hf. Every atom of ^{182}Hf decays to a daughter atom of ^{182}W; therefore,

$$\left(^{182}\mathrm{W}\right)_{\mathrm{today}} = \left(^{182}\mathrm{W}\right)_{\mathrm{original}} + \left(^{182}\mathrm{Hf}\right)_{\mathrm{original}} \quad [1]$$

Because it is easier to measure these effects using isotopic ratios rather than absolute numbers of atoms, we divide by another isotope of W:

$$\left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}}\right)_{\mathrm{today}} = \left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}}\right)_{\mathrm{original}} + \left(\frac{^{182}\mathrm{Hf}}{^{184}\mathrm{W}}\right)_{\mathrm{original}} \quad [2]$$

However, the ^{182}Hf is no longer extant and so cannot be measured. For this reason, we convert eqn [2] to a form that includes a monitor of the amount of ^{182}Hf that would have been present as determined from the amount of Hf today. Hafnium has several stable nuclides and ^{180}Hf is usually deployed. Hence, eqn [2] becomes

$$\left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}}\right)_{\mathrm{today}} = \left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}}\right)_{\mathrm{original}} + \left\{\left(\frac{^{182}\mathrm{Hf}}{^{180}\mathrm{Hf}}\right)_{\mathrm{original}} \times \left(\frac{^{180}\mathrm{Hf}}{^{184}\mathrm{W}}\right)_{\mathrm{today}}\right\} \quad [3]$$

which represents the equation for a straight line. A plot of ^{182}W/^{184}W against ^{180}Hf/^{184}W for a suite of cogenetic samples or minerals will define a straight line, the slope of which gives the ^{182}Hf/^{180}Hf at the time the object formed. This can be related in time to the start of the solar system with Soddy and Rutherford's equation for radioactive decay:

$$\left(\frac{^{182}\mathrm{Hf}}{^{180}\mathrm{Hf}}\right)_{\mathrm{original}} = \left(\frac{^{182}\mathrm{Hf}}{^{180}\mathrm{Hf}}\right)_{\mathrm{BSSI}} \times e^{-\lambda t} \quad [4]$$

in which *BSSI* is the bulk solar system initial ratio, λ is the decay constant (or probability of decay in unit time), and t is the time that elapsed since the start of the solar system. Using this method and the $(^{182}\mathrm{Hf}/^{180}\mathrm{Hf})_{\mathrm{BSSI}}$ of $\sim 1 \times 10^{-4}$ (Table 1), it has been possible to determine the age of a wide range of early solar system objects. Note that some early papers utilized a $(^{182}\mathrm{Hf}/^{180}\mathrm{Hf})_{\mathrm{BSSI}}$ of $\sim 1 \times 10^{-5}$ (Jacobsen and Harper, 1996). Later, papers used a value of $\sim 2 \times 10^{-4}$ (e.g., Lee and Halliday, 1997). Now, it is clear that the value must be $\sim 1 \times 10^{-4}$ (Kleine et al., 2002; Schönberg et al., 2002; Yin et al., 2002).

There have been several reviews of ^{182}Hf–^{182}W chronometry, none of which is complete because the field is changing fast. The most recent comprehensive modern review is that provided by Kleine et al. (2009). The principle of the technique is that Hf/W ratios are strongly fractionated by core formation because W normally is moderately siderophile, whereas Hf is lithophile (Figure 12). Therefore, in the simplest of models, the W isotopic composition of a silicate reservoir such as the Earth's PM is a function of the timing of core formation (Figure 13(a)). This could be determined from eqns [3] to [4] that together provide the time (usually referred to as a two-stage model age) when the Hf/W ratio was fractionated by core formation.

For the ^{182}Hf–^{182}W system, this time is given as

$$t_{\mathrm{CHUR}} = \frac{1}{\lambda} \ln \left[\left(\frac{^{182}\mathrm{Hf}}{^{180}\mathrm{Hf}} \right)_{\mathrm{BSSI}} \times \left(\frac{\left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}} \right)_{\mathrm{SAMPLE}} - \left(\frac{^{182}\mathrm{W}}{^{184}\mathrm{W}} \right)_{\mathrm{CHONDRITES}}}{\left(\frac{^{180}\mathrm{Hf}}{^{184}\mathrm{W}} \right)_{\mathrm{SAMPLE}} - \left(\frac{^{180}\mathrm{Hf}}{^{184}\mathrm{W}} \right)_{\mathrm{CHONDRITES}}} \right) \right] \quad [5]$$

where t_{CHUR} is the time of separation from a CHondritic Uniform Reservoir. If, however, a planet grows over tens of millions of years and the core grows as the planet gets larger, as is nowadays assumed to be the case for the Earth (**Figure 13(b)**), the W isotopic composition of the PM is also a function of the rate of growth or the longevity of accretion and core formation. In this case, the equations need to be more complex (Jacobsen, 2005) or numerical models need to be established that calculate the W isotopic effects as a result of each stage of accretion or growth (Halliday, 2004).

As already stated, the earliest precisely dated objects that are believed to have formed within our solar system are the CAIs found in many chondrites (**Figure 2**). It would appear that at least some chondrules formed a few million years after the start of the solar system. Therefore, isotopic dating provides evidence that the disk from which the planets grew was still a dusty environment with rapid melting events, say, 2 million years after the start of the solar system (**Figure 2**). This means that chondrites must also have formed after this time.

Differentiated asteroids and their planetary cores and silicate reservoirs, as represented by some iron and basaltic achondrite meteorites formed *before* chondrites. A variety of chronometers have been used to demonstrate that iron meteorite parent bodies, the eucrite parent body, and the angrite parent body appear to have formed in the first few million years of the solar system (**Figure 2**) (Kleine et al., 2005a,b; Markowski et al., 2006; Scherstén et al., 2006). With extensive replication and better mass spectrometers, very high precision can now be achieved on the W isotopic compositions of iron meteorites. The latest data for iron meteorites provide evidence that accretion and core formation were very short-lived (Kleine et al., 2005a,b; Kruijer et al., 2012; Markowski et al., 2006; Qin et al., 2008; Scherstén et al., 2006). By correcting for cosmogenic effects, it can be demonstrated that some magmatic iron meteorites, thought to represent planetesimal cores, formed within 2 million years of the start of the solar system. Therefore, they appear to represent examples of early planetary embryos, as predicted from dynamic theory.

In a similar manner to differentiated asteroids, Mars also seems to have accreted quickly. The most recent compilations of data for martian meteorites (Foley et al., 2005; Halliday and Kleine, 2006; Kleine et al., 2004a,b; Lee and Halliday, 1997) provide a clear separation of W isotopic compositions between nakhlites and shergottites and confirm the view that accretion and core formation on Mars were fast. Some recent models (Dauphas and Pourmand, 2011; Halliday and Kleine, 2006) place the timescale for formation of Mars at less than one million years (**Figure 2**). If this is correct and the ages *are* based on assumptions about the model Hf/W of the martian mantle, Mars probably formed by a mechanism such as runaway growth, rather than by protracted collision-dominated growth. In other words, Mars may represent a unique example of a large primitive planetary embryo with a totally different accretion history from that of the Earth.

The conclusion from these studies of short-lived isotopic systems in meteorites is that chondrites, though widely considered the most primitive objects in the solar system and the building blocks for the planets, actually did not form for at least 2 million years, longer than the timescales for growth of planetary embryos determined theoretically. Where and how they formed are currently unclear. The chronometry is consistent with the chemistry in indicating that chondrites do not represent the actual building blocks of the Earth. In fact, it has long been thought that the Earth's growth was dominated by accretion of differentiated objects (Taylor and Norman, 1990). Chondrites almost certainly represent a good approximation of the average disk material but models that use chondrite variability to extrapolate to the bulk composition of the Earth, or attempt to build the Earth from chondritic components, need to be treated with some awareness of this issue.

9.02.10 Rates of the Earth's Accretion and Differentiation

Although Patterson determined the age of the Earth, there is in fact no such thing as a precise time when all of the constituents were amalgamated. It is more reasonable to instead assume a style of growth and use the isotopic data to define a rate of change of planetary growth from the integrated history of the chemical fractionation of the parent/daughter ratio. The latter is provided by the isotopic composition of the daughter.

The most powerful chronometers for estimating the growth rate of the Earth are ^{182}Hf–^{182}W and ^{235,238}U–^{207,206}Pb (Allègre et al., 1995a; Galer and Goldstein, 1996; Halliday, 2004, 2014). In both cases, the parent/daughter ratio is strongly fractionated during core formation (**Figure 12**). As such, the W and Pb isotopic compositions in the silicate portion of the planet reflect how fast it grew and internally fractionated by metal segregation. Although it was once thought that accretion was rapid (Hanks and Anderson, 1969) and formation of the core was protracted (Solomon, 1979), there exists strong evidence that the process of separating core material from mantle material and delivering it to the core after it arrives by impact is a faster one (Sasaki and Nakazawa, 1986) than the process of accretion itself; this is discussed in **Chapter 9.03** by Rubie et al. As such, the core grows proportionally, limited in size by the growth of the planet. On this basis, Pb and W isotopic data yield constraints on the rates of growth of the planet, provided certain assumptions are valid.

In practice, the ^{182}Hf–^{182}W system provides a more powerful constraint on how soon accretion and core formation started, whereas ^{235,238}U–^{207,206}Pb provides a better indication of how long it could have persisted. The ^{182}Hf–^{182}W system is better constrained in terms of the values of key parameters in Earth. To determine accretion rates, one needs to know the parent/daughter ratio in the total planet and in the reservoir, the formation of which is being dated. This is sometimes complex because we do not know independently how much resides in the inaccessible core. Some elements like Hf and W are refractory and were not affected by heating in the inner portions of the disk. Therefore, the Hf/W ratio of the total Earth is well known from independent measurements of chondrites (Newsom, 1995), even if the Earth is not exactly chondritic. However, Pb is moderately volatile and therefore variably depleted in the Earth, Moon, Vesta, and Mars and almost certainly in Venus and Mercury. Therefore, two distinct

processes have produced depletions of Pb in the silicate Earth – core formation and volatile loss. Two different methods are used to estimate the amount of Pb in the total Earth. The first is a comparison between the amount of Pb depletion in chondrites and that of other volatile elements that are not partitioned into planetary cores (Allègre et al., 1995a) (**Figure 6**). The second is based on estimates of the amount of volatile depletion expected given the average condensation temperature (Galer and Goldstein, 1996) (**Figure 7**). These two approaches are in broad agreement for Pb.

In addition to parent/daughter ratio, isotopic methods require a well-defined value for the isotopic composition of the silicate Earth relative to the rest of the solar system. The W isotopic composition of the silicate Earth is indeed extremely well defined because after 4.5 billions of convective mixing, the Earth's silicate reservoir has homogenized early isotopic variability. The W isotopic age of the Earth could in principle be determined from the composition of a tungsten carbide drill bit. However, with W, the problem has been to get the correct value for the average solar system. The earliest Hf–W papers assumed a certain value (Jacobsen and Harper, 1996). Early measurements of chondrites were variable (Lee and Halliday, 1996, 2000a,b). Although most of the early measurements for iron meteorites, achondrites, lunar samples, martian meteorites, and enstatite chondrites have reproduced well with more modern high-precision methodologies, it is now recognized that the early W isotopic measurements for carbonaceous chondrites in particular, reported in Lee and Halliday (1996), are incorrect and the timescales have had to be reassessed in the light of this (Halliday, 2004; Kleine et al., 2002; Schönberg et al., 2002; Yin et al., 2002). The problem for U–Pb is almost the opposite. The present-day value for the silicate Earth is very hard to constrain. The reason is that ^{235}U and ^{238}U are still alive in the Earth and capable of producing isotopic variations in ^{207}Pb/^{204}Pb and ^{206}Pb/^{204}Pb. Therefore, the Earth is very heterogeneous in its Pb isotopic composition (Galer and Goldstein, 1996). This is shown in **Figure 11** in which 11 estimates of the average Pb isotopic composition of the silicate Earth are presented. Each of these would imply a different timescale for the Earth's accretion, as discussed later in the text and shown in **Figure 11**.

The simplest kind of age calculation derives a time of fractionation or two-stage model age by assuming that the fractionating process was instantaneous (**Figure 13(a)**). This is the equivalent of assuming that the core formed instantaneously from a fully formed Earth. The ^{182}Hf–^{182}W model age so determined is ~30 Ma (Kleine et al., 2002; Schönberg et al., 2002; Yin et al., 2002), whereas the ^{235,238}U–^{207,206}Pb ages are more protracted (**Figure 11**). The possible reasons for this are discussed in more detail in the succeeding text. More complex models calculate an accretion rate (Halliday, 2000, 2004; Jacobsen and Harper, 1996; Kleine et al., 2009) by making assumptions about the style of growth. Generally speaking, this is an exponentially decreasing rate of growth that emulates what has arisen from dynamic simulations (**Figure 14**). This is consistent with the kinds of curves produced by George Wetherill who modeled the growth of the terrestrial planets using Monte Carlo simulations. To quantify the accretion timescale with an exponentially decreasing rate of growth requires the use of a time constant for accretion as follows:

$$F = 1 - e^{-(1/\tau)\times t} \qquad [6]$$

where F is the mass fraction of the Earth that has accumulated, τ is the mean life for accretion (or inverse of the time constant), and t is the time (in the same units). The mean life corresponds to the time taken to achieve 63% growth. The accretionary mean life for ^{182}Hf–^{182}W is ~11 Ma (Yin et al., 2002) but is again much longer for ^{235,238}U–^{207,206}Pb (Halliday, 2004, 2008, 2014) (**Figure 15**). Some of the most recent models

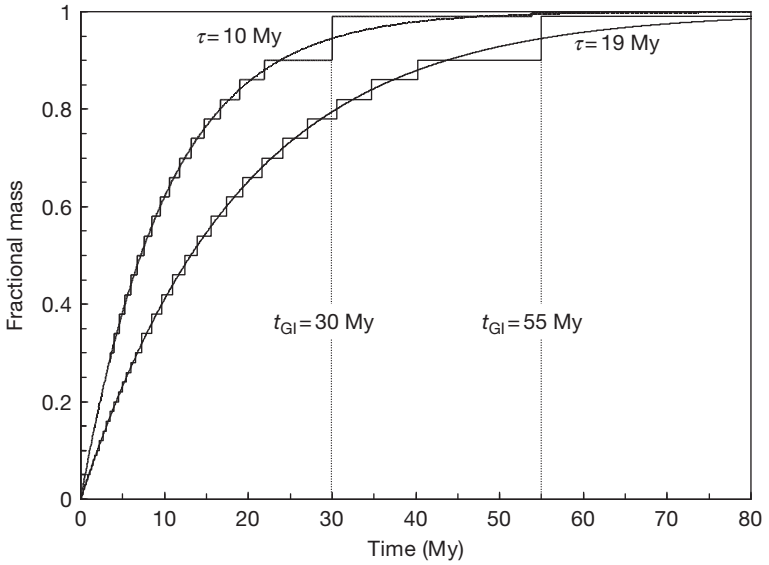


Figure 14 The mean life of accretion of the Earth (τ) is the inverse of the time constant for exponentially decreasing oligarchic growth from stochastic collisions between planetary embryos and planets. The growth curves corresponding to several such mean lives are shown including the one that most closely matches the calculation made by the late George Wetherill based on Monte Carlo simulations. The mean life determined from tungsten isotopes is in excellent agreement with Wetherill's predictions. Reproduced from Halliday AN (2004) Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* 427: 505–509.

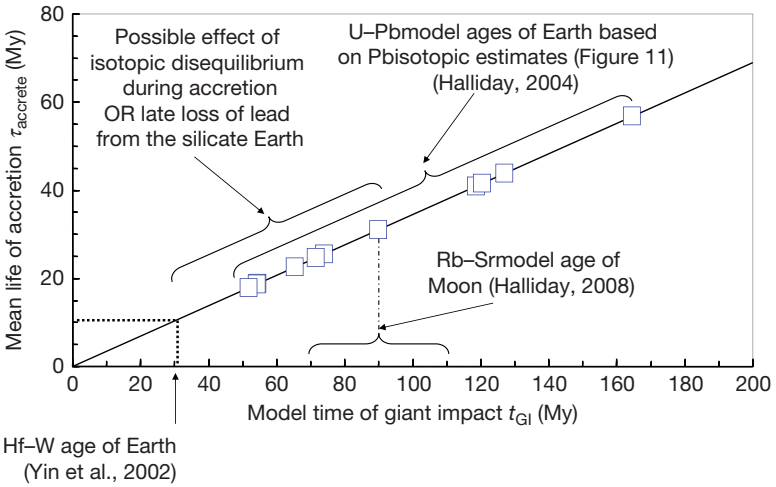


Figure 15 Calculated values for the Earth's mean life of accretion (τ) and time of the giant impact (t_{GI}) given in million years ago as deduced from the different estimates of the Pb isotopic composition of the BSE (**Figure 11**) using the type of accretion model shown in **Figure 14** (see Halliday, 2004). Note that while all calculations assume continuous core formation and total equilibration between accreted material and the BSE, the accretion is punctuated, as predicted from the planetesimal theory of planetary accretion. This generates more protracted calculated timescales than those that assume smooth accretion (Halliday, 2004). The $^{238}U/^{204}Pb$ values assumed for the Earth = 0.7 (Allègre et al., 1995a,b). Even the Hf–W timescales using the exact same style of model are significantly shorter.

(Halliday, 2004) also try to simulate the effect of a more stochastic style of accretion with growth from sporadic large (Moon- to Mars-sized) impactors (**Figure 14**). The staggered growth model shown in **Figure 14** assumes that the Earth grew from 90% to 99% of its current mass as a result of a collision with an impacting Mars-sized planet called 'Theia' that produced the Moon. The changes in the rate of accretion with time are calculated from the timing of this event. Therefore, the model time of the giant impact is directly related to the mean life of accretion in these models (**Figure 15**). Regardless of the details of the accretion processes, this last major impact would have had a dominant effect on the present-day W isotopic composition of the BSE. Therefore, predictions about the last major accretion event to affect the Earth, which are based on the W isotopic composition of the BSE, can in principle be tested by determining the age of the Moon. However, this has been less than straightforward.

9.02.11 The Age of the Moon

Despite considerable efforts using long-lived chronometers, the age of the Moon was for decades only known to be some time in the first 150 Ma of the solar system (Wasserburg et al., 1977). With the advent of Hf–W chronometry, it has been possible to determine that the Moon must have formed tens of millions of years after the start of the solar system. This is because the Moon has a high Hf/W of ~25 (König et al., 2011), and although it is thought to have formed in a single event, it has relatively unradiogenic W just like the Earth (Touboul et al., 2007). This means the Moon cannot have formed before about 37 Ma (Halliday, 2014). The first attempts to date the Moon with Hf–W (Lee et al., 1997) revealed some samples with relatively high ^{182}W in many lunar samples and the fact that these compositions were variable provided evidence that there was still live ^{182}Hf in the Moon during its earliest differentiation. Subsequently, it was shown that in some of these, a portion of the effect was due to cosmogenic production of ^{182}Ta that decays to ^{182}W (Leya et al., 2000). This can be corrected in some cases by using exposure ages or internal mineral systematics (Lee et al., 2002). A novel and in many respects more straightforward and reliable approach was proposed by Kleine et al. (2005a,b) who used carefully separated metals with low Ta/W. All of these approaches yield similar timescales of between about 30 and 55 Ma after the start of the solar system. However, subsequent work by Touboul et al. (2007) showed that no effects from the decay of ^{182}Hf could be reliably resolved; all appeared to be cosmogenic. Since the publication of Touboul et al.'s paper, the best estimate of the Hf/W of lunar and terrestrial mantles has been shown to be remarkably similar at ~25 (König et al., 2011). The identical W isotopic composition and Hf/W provide powerful evidence that the Moon's atoms equilibrated with or were derived from those of the Earth. However, the age constraint is limited to ~37 Ma at the earliest (Halliday, 2014). This still provides strong support for the giant impact theory since such a late origin for an object of the size of the Moon is not readily explicable unless it is derived from a previously formed planet.

There are other isotopic systems that have been used to constrain the age of the Moon. In particular, the Rb–Sr system was deployed Carlson and Lugmair (1988) to define the age of the Moon assuming that its initial Sr isotopic composition was derived from the Earth, the Rb/Sr ratio (0.03) for which is well established. Initially, this approach appeared flawed because dynamic simulations of the giant impact derived most of the mass of the Moon from the impactor planet Theia (Canup and Asphaug, 2001). However, the average oxygen isotopic compositions of lunar samples cannot be resolved from the terrestrial fractionation line at the 5 ppm level (Wiechert et al., 2001). Other isotopic systems that are identical include Si (Armytage et al., 2012), Ti (Leya et al., 2008; Zhang et al., 2012), Cr (Lugmair and Shukolyukov, 1998), and W (Touboul et al., 2007). Whereas the O, Ti, and Cr isotopic effects could reflect formation from a similar inner solar system

provenance (Wiechert et al., 2001), the Si and W isotopic compositions reflect the history of core formation and are more likely to be planet-specific. Therefore, the atoms of the Moon either were derived from the Earth (Zhang et al., 2012) or reflect a process of large-scale equilibration between the Moon and Earth (Pahlevan and Stevenson, 2007; Pahlevan et al., 2011). Either way, the Rb/Sr of the Earth can be used in combination with the Moon's initial Sr isotopic composition to derive an age of 90 ± 20 Ma (Halliday, 2008). It has been proposed that the Earth's Rb/Sr is in fact lower because of impact erosion (O'Neill and Palme, 2008). In this case, the age is 125 ± 25 Ma (Halliday, 2014).

One other approach to defining the age of the Moon is to place a minimum age constraint from the ages of the oldest precisely dated lunar rocks. For a long time, this was given by the U–Pb age of 70 ± 10 Ma for lunar ferroan anorthosite 60025 (Hanan and Tilton, 1987). This ferroan anorthosite has been redated by Borg et al. (2011) and now appears to be much younger and closer to 200 Ma. Therefore, while 90 ± 20 Ma is shown as the Rb–Sr model age of the Moon in Figures 2 and 15, in reality, there is currently considerable uncertainty about the exact age. It does lie in the range 30–200 Ma and overlaps ages for the oldest terrestrial zircons (Wilde et al., 2001).

9.02.12 First Principles of Chemical Constraints on Core Formation

Having used the ^{182}Hf–^{182}W system to constrain the timing of core formation, the physical conditions of the core separation process can be estimated, provided chemical compositions of the BSE, the core, and the bulk Earth are known. As discussed earlier, the BSE exhibits a depletion in volatile elements with respect to the chondritic reference model. It also shows depletions in refractory siderophile elements such as W, Mo, Re, and Os. In these latter cases, the depletions cannot be due to volatility and must be the result of partial extraction into the core. The extent of extraction can be estimated by comparing silicate Earth concentrations (normalized to CI chondrites) to those of refractory lithophile elements such as the REE. This enables us to calculate, for each element, a core–mantle partition coefficient D_i defined as follows:

$$D_i = [i]_{\text{core}} / [i]_{\text{silicate Earth}} \quad [7]$$

where $[i]$ is the concentration of element i. Lithophile elements have D values close to zero. Table 2 shows that siderophile elements exhibit a very wide range of core–mantle partitioning behavior reflecting their different chemical properties and the conditions under which core segregation took place.

Core–mantle partitioning is best defined for those refractory elements that are weakly or moderately siderophile and that, like Ni and Co, are compatible in solid mantle silicates. The concentrations of these elements vary little in mantle samples, which means that D_i values based on the chondrite reference model have small uncertainties. Highly siderophile refractory elements such as the Pt group (Figure 7) are slightly less well constrained. The abundances of these elements in mantle samples are very low and rather variable. It is generally accepted, however, that their ratios one to another in the silicate Earth are, as shown in Figure 7, approximately chondritic.

Table 2 Core–mantle partition coefficients

Element	McDonough and Sun (1995) and McDonough (2003)	Allègre et al. (1995a,b)	Likely range	Low pressure Experimental D_i (Newsom, 1990; O'Neill and Palme, 1998)
D_{Fe}	13.66	13.65	13.65	13.65
D_{Ni}	26.5	24.4	23–27	4900
D_{Co}	23.8	24.7	23–27	680
D_{V}	1.83	–	1.5–2.2	0.02
D_{W}	16	–	15–22	3
D_{Pd}	800	–	600–1000	7×10^5
D_{Ir}	800	–	600–1000	10^{11}
D_{Pt}	800	–	600–1000	4×10^6
D_{Nb}	–	–	0.2–0.8	–
D_{Cr}	3.4	2.9	0.5–3.5[a]	0.2
D_{Mn}	0.29	5	0.2–2.0[a]	0.006
D_{Si}	0.29	0.34	0.1–0.35[a]	10^{-5}
D_{S}	76	–	50–100[a]	50
D_{Ga}	–	–	0–1.5[a]	15
D_{P}	22	45	20–50[a]	30

[a]Large uncertainty due to volatility.

There are a number of volatile siderophile elements such as S, Se, Te, P, and possibly Si whose core concentrations are difficult to estimate because their bulk Earth contents are not well constrained. This means, for example, that depletions in Mn, Si, and Cr (Figure 7) could be due to volatility or to sequestration in the core or both. The lack of constraint means that the uncertainties in D_i are large.

Segregation of the reduced core from the oxidized mantle took place at high temperatures, and since transfer from mantle to core involves reduction, element partitioning between the two must have depended on oxygen fugacity. From the current concentration of FeO in the mantle of about 8 wt% and of Fe in the core of 85% (Allègre et al., 1995b; McDonough, 2003; McDonough and Sun, 1995), it is fairly straightforward to estimate that core segregation took place at ~ 2 logfO_2 units below Fe–FeO (IW) equilibrium (e.g., Li and Agee, 1996).

Table 2 shows a comparison of the calculated core(metal)–mantle(silicate) partition coefficients and those obtained experimentally at low pressures (0–2GPa), high temperatures ($\sim$1800 K), and oxygen fugacity corresponding to 8 wt% FeO in the mantle (Newsom, 1990; O'Neill and Palme, 1998). As can be seen, the observed core–mantle partition coefficients are, for many of the siderophile elements, much (orders of magnitude) smaller than those determined experimentally. This manifests itself in the concentrations of many elements of concern (e.g., Ni, Co, and Pd) being much greater in the mantle than would be predicted from the experimental data, the so-called excess siderophile problem. Of equal importance are the weakly siderophile elements such as V and Cr for which the mantle concentrations are *lower* than predicted from the experimental data. Clearly, the mantle concentrations of these two groups of elements cannot be explained by equilibrium with metallic core at the fixed pressure, temperature, and oxygen fugacity conditions of Table 2. Early core segregation in small planetesimals

and planetary embryos, including proto-Earth, should, however, have resulted in siderophile element abundances generally consistent with the *D* values given in Table 2. The Earth, apparently, has inherited little of the geochemistry of these earlier events.

9.02.13 Explanations for the 'Excess Siderophile Problem'

Given that the 'excess siderophile problem' was not inherited, but arose from the manner of core formation on the Earth, several possible explanations have been proposed:

(1) *Inefficient core formation* (Arculus and Delano, 1981; Jones and Drake, 1986). In this model, some fraction of the core was left behind in the mantle, together with its siderophile element budget. Later, reoxidation and remixing of the metal led to the observed elevated mantle concentrations of siderophile elements. Principal objections to this model are that it can only explain the mantle abundances of some elements and that an enormous amount of water, about 15 times the current mass of the hydrosphere, would be required to reoxidize the 'retained' metal (O'Neill, 1991a,b).

(2) *Heterogeneous accretion* (Chou, 1978; O'Neill, 1991b; Wänke, 1981). In this model, it is suggested that the conditions of core segregation changed from reducing at the beginning of accretion to more oxidized toward the end. During the earliest, reduced, phase, highly and moderately siderophile elements (e.g., Ni, Co, and Pt group) would have been completely extracted to the core, together with some fraction of the weak siderophiles such as Si, V, and Cr. As conditions became more oxidizing, core extraction of siderophile elements would have ceased progressively, beginning with the weakest siderophiles and ending with more strongly metallic elements. The final phase, after core formation had ceased, was the addition of a 'late veneer' of about 0.5% of chondritic material. The latter raised the concentrations of highly siderophile elements in the mantle and, since core formation had ceased, ensured their chondritic proportions one to another (Figure 7). This model solves the 'excess siderophile problem,' but, being a disequilibrium model with multiple steps, is very difficult to test.

(3) *High-pressure core formation*. Murthy (1991) proposed that the partition coefficients for siderophile elements between metal and silicate should approach unity at high temperatures. This led to a variety of important new studies, and recent measurements of metal–silicate partitioning at high pressures and temperatures (Gessmann and Rubie, 1998; Li and Agee, 1996, 2001; Righter and Drake, 2000; Righter, et al., 1997) have shown that the partition coefficients of some siderophile elements change such that their mantle abundances, inexplicable by low-pressure equilibrium, may be explained by metal–silicate equilibrium at very high pressures and temperatures (e.g., Li and Agee, 1996; Righter and Drake, 1997). These observations led to the 'deep magma ocean' hypothesis to explain the 'excess siderophile' problem. As the Earth grew, it is argued (e.g., Righter and Drake, 1999) that droplets of metallic liquid descended through a 700–1200 km deep (28–40 GPa) magma ocean, ponding at its base in equilibrium with the magma ocean. The liquid metal subsequently descended in large diapirs to the growing core without further reaction with the surrounding silicate (Figure 16). High-pressure core formation can explain, by metal–silicate equilibrium, the partitioning of many elements between the core and mantle. It does still require, however, the 'late veneer' of chondritic material to explain the concentrations and proportions of the highly siderophile elements in the silicate Earth.

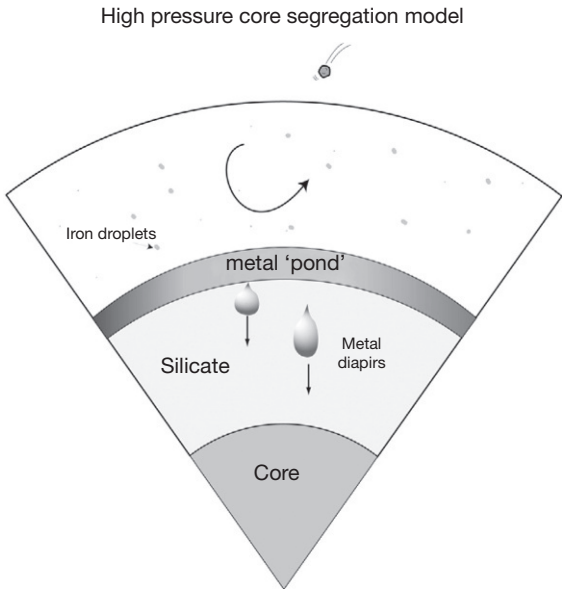


Figure 16 Core formation in the 'deep magma ocean' model. Impacting planetesimals disaggregate and their metallic cores break into small droplets due to Rayleigh–Taylor instabilities. These droplets descend slowly, reequilibrating with the silicate until they reach a region of high viscosity (solid), where they pond in a layer. The growing dense metal layer eventually becomes unstable and breaks into large blobs (diapirs) that descend to the core without further interaction with the silicate. The liquidus temperature of the silicate mantle should correspond to pressure and temperature conditions at a depth above the lower solid layer and plausibly within the metal layer.

9.02.14 The 'Deep Magma Ocean' Model of Core Formation

As the Earth grew, the gravitational energy deposited by accreting planetesimals would have increased, and at about 10% of its current size (Stevenson, 1981) was sufficient for significant melting to occur. After this point, the Earth would have periodically had an extensively molten outer layer (a magma ocean) of variable thickness. In some cases, the impact energy would have been sufficient to melt both the impactor and the proto-Earth (Cameron and Benz, 1991; Canup and Asphaug, 2001). From W isotopic evidence, it is clear that planetary cores segregated from their silicate mantles very early. As a result, most incoming planetesimals and planets are likely to have been already differentiated into silicate and metal like Vesta, Mars, and the Moon-forming planet Theia (Taylor and Norman, 1990; Yoshino et al., 2003). The fates of preexisting cores would have been dependent on the sizes of the Earth and impactor. Continuous core formation isotopic models normally assume that all of the metal

and silicate in the impactor mixes and equilibrates with the silicate portion of the Earth before further growth of the Earth's core (Kleine et al., 2002; Schönberg et al., 2002; Yin et al., 2002). However, the degree to which the impactor's core fragments depends on the angle of impact and relative size of the bodies (see discussion by Stevenson and in Rubie et al., this volume). Large fragments of metal could, under some circumstances, have mixed directly with the Earth's core (Dahl and Stevenson, 2010; Halliday, 2004; Karato and Rama Murthy, 1997) (**Figure 13(c)**). However, the size of those droplets depends on the fluid dynamics of the process.

Rubie et al. calculated the size of the droplets of liquid metal raining out of a magma ocean and concluded that these should have been about 1 cm diameter. Droplets of this small size would reequilibrate rapidly with the silicate melt as they fell (Rubie et al., 2003). Liquid metal and silicate would therefore have continued to reequilibrate until the former either reached the core–mantle boundary (if the mantle were completely molten) or collected at a level above a solid, high-viscosity lower layer (**Figure 15**). In the latter case, the metal layer would cease reequilibrating after it had reached about 5 km in thickness (Rubie et al., 2003) and because of the high viscosity of the lower layer would segregate as large diapirs to the core (Karato and Rama Murthy, 1997). The droplet model can apply in the case of small bodies accreting to the growing Earth, but may not apply after a giant impact such as the Moon-forming impact (see **Chapter 9.03**). In the latter case, isotopic and chemical equilibration of metal from the impactor with the silicate portion of the Earth depends on the extent to which the core of the impactor is fragmented.

The fact that the silicate Earth has a W isotopic composition that is so similar to average solar system (chondrites) means that in general terms, metal–silicate equilibration must have been the norm. However, the small difference that is now known to exist between the W isotopic compositions of chondrites and the BSE could reflect an event like the giant impact that involved particularly rapid amalgamation of large fragments of impactor metal with the Earth's core with little chance for equilibration with the silicate portion of the Earth (Halliday, 2004).

If we know the age of the Moon, we know the timing of the last accretion stage in the history of the Earth (**Figures 2** and **15**). From this, we can determine the extent to which the W isotopic composition of the silicate Earth reflects planetary-scale disequilibrium of added material during accretion. Preliminary attempts at calculating this (Halliday, 2004; Halliday and Kleine, 2006; Kleine et al., 2004b) showed that on average, the degree of equilibration between incoming metal and the Earth's silicate must have been high (70–90%). However, with a late (say 70–200 Ma) giant impact, the W isotopic compositions become harder to explain (Halliday, 2008). For example, with a 90 Ma giant impact and an exponentially decreasing stepped accretion rate (Halliday, 2004) (**Figure 14**), the W isotopic composition of the BSE is consistent with 55% of the cores of impactors merging directly with the Earth's core (**Figure 13(c)**). This might be more easily reconciled with dynamic simulations. However, it should be emphasized that greater resolution of the simulations and a more comprehensive understanding of the physics of metal–silicate mixing under such extreme conditions are needed before one can say which isotopic model is more consistent with the likely physical processes.

In the event that the energetics of impact and core segregation lead to a completely molten mantle, numerical simulations (Abe, 1997; Solomatov, 2000, **Chapter 9.04**) indicate that such a deep magma ocean would be short-lived and that the lower mantle would crystallize in a few thousand years. A shallower partially molten layer would crystallize much more slowly, however, and could remain as a mixture of crystals and melt for 100 Ma (Abe, 1997; Solomatov, 2000, **Chapter 9.04**). The energetics of impact, core segregation, and mantle crystallization leads to a dynamic view of the growing Earth in which the outer molten part deepened and shallowed many times after episodic impact. The pressures and temperatures recorded by core–mantle partitioning are therefore values averaged over numerous cycles of metal accumulation and segregation such as that depicted in **Figure 16**.

If one considers the magma ocean model of **Figure 16** and takes account of the likelihood that the magma ocean shallowed and deepened many times during accretion and core segregation, several principles emerge. Firstly, accumulation of metal at the base of a completely or partially molten silicate layer, as shown in **Figure 15**, implies temperatures close to or slightly below the liquidus of the mantle peridotite. Secondly, average pressure should have increased as the planet grew. Thirdly, since the silicate liquidus has a positive pressure–temperature slope, average temperature of core segregation should also have increased as the planet grew. Having accepted these principles, we can use the dependences of D_i on pressure and temperature (Chabot and Agee, 2003; Li and Agee, 2001; Righter et al., 1997; Wade and Wood, 2005) to develop constraints on the core segregation process.

9.02.15 Core Segregation During the Growth of the Earth

If we consider the case of instantaneous core formation (**Figure 13(a)**), then current partitioning data for the elements of **Table 2** yield average pressures and temperatures of ~40 GPa and ~3750 K, respectively (Wade and Wood, 2005). Given that the mantle has superchondritic Mg/Si, these conditions imply that a substantial amount of Si, more than required to give the bulk Earth a chondritic Mg/Si ratio, resides in the core (Wade and Wood, 2005). The reason is that Si partitioning into the metal is strongly temperature-dependent and the calculated temperature is very high, in fact 650° above the silicate liquidus. This seems unlikely because the base of the magma ocean should be at or below the mantle liquidus (**Figure 16**).

Continuous accretion and core formation models (**Figure 13(b)**) do not change the result, temperatures remaining above the silicate liquidus. Apart from invoking wholesale disequilibrium, the simplest solution (Wade and Wood, 2005), as illustrated in **Figure 17**, is to allow oxygen fugacity (as represented by the FeO content of the mantle) to increase as accretion and core segregation progress.

Figure 17 illustrates a continuous process of accretion and core segregation in 1% intervals. New material added to the Earth was assumed to be mixed with the preexisting mantle, but isolated from the protocore. The core was segregated in 1% steps, each aliquot being in equilibrium with the entire mantle at the time of segregation at pressures constrained by Ni and Co partitioning and temperatures on the peridotite liquidus

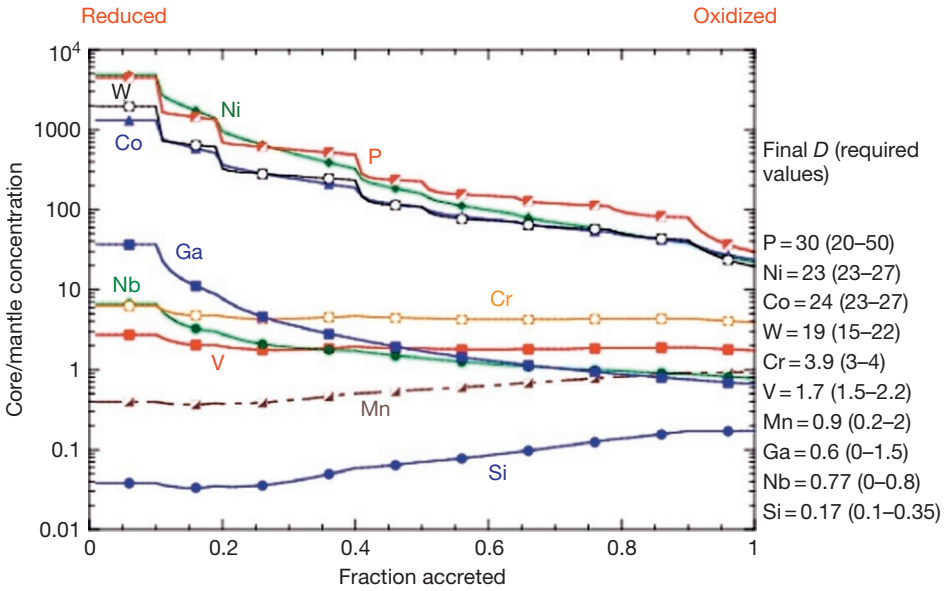


Figure 17 Figure showing one possible path of accretion and core segregation for the Earth. The Earth was accreted and the core segregated in 1% intervals. At each step, the silicate Earth was homogenized and the core segregated at the base of the magma ocean of **Figure 16**. At 100% accretion (right-hand side), the core/mantle partitioning of each element must match the 'required value' of **Table 2**. In order to achieve this and maintain temperatures on the silicate liquidus, the Earth must become more oxidized during accretion. In the example shown, the oxygen fugacity increases in small steps from about 3.5 log units below the Fe–FeO (IW) buffer to 2 log units below IW during accretion and the magma ocean depth corresponds to 30–40% of the depth to the core–mantle boundary in the growing planet.

(Wade and Wood, 2005). The latter were found to be consistent with magma oceans extending to 30–40% of the depth of the growing mantle. A large number of oxygen fugacity paths yield the partitioning of V, W, Nb, Ni, Co, Mn, Cr, Ga, P, and Si, which are consistent with the estimates of **Figure 17**. All require, however, that most of the Earth formed under conditions more reducing than those implied by the current FeO content of the mantle and that the Earth became more oxidized toward the end of accretion. The combined effects of increasing temperature and pressure during accretion lead to a progressive increase in core Si content (**Figure 17**) such that the final core contains about 4% Si. This brings the Mg/Si ratio of the bulk Earth much closer to chondritic than implied in **Figure 8**. As required by all accretion models, 0.5% of chondritic material must be added after cessation of core formation in order to generate the chondritic ratios of the highly siderophile elements such as Pd, Ir, and Pt (**Figure 17**). The current 'best explanation' is therefore an amalgam of 'deep magma ocean' and 'heterogeneous accretion' hypotheses, that is, core segregation at the base of a deepening magma ocean under progressively more oxidizing conditions (Wade and Wood, 2005).

The idea that the Earth became more oxidized during accretion has been proposed on a number of occasions (e.g., O'Neill, 1991a,b), but this leads to two other important questions. Firstly, the Earth has a substantially higher core/mantle mass ratio than Mars and is hence, as a planet, more reduced than this smaller body. What caused the Earth to be more reduced than smaller bodies? Was it due to part of the silicate Earth being removed during accretion (O'Neill and Palme, 2008) as proposed for Mercury (Benz et al., 1987)? If the Earth later became more oxidized, as evidenced by the relatively oxidized nature of terrestrial as opposed to lunar or martian basalts (Karner, et al., 2006) and the constraints of core–mantle partitioning discussed earlier in the text, then what was the mechanism of oxidation and when did it operate?

9.02.16 Oxidation State of the Earth During and After Accretion

As planetesimals accreted during the first few million years of solar system history, they would have been surrounded by a hydrogen-rich nebular gas of solar composition. This gas would have had little effect on the chemical compositions of small bodies because the pressure was very low (10^{-4} bar). As protoplanets grew, however, their masses might have become high enough for them to capture a nebular gas atmosphere (Hayashi et al., 1979, 1985; Ikoma and Genda, 2006; Sasaki and Nakazawa, 1986) such that at a mass >20% of the current mass of Earth hydrogen pressures >1 bar are plausible.

There is a problem with these models because by the time a body that is sufficiently massive to retain a large hydrogen atmosphere has formed, the nebula may have already dissipated. Ikoma and Genda (2006) pointed out that the atmosphere thereby accreted is only logarithmically sensitive to the nebula density, which implies that a massive hydrogen atmosphere can be accreted even when the nebula has largely departed. However, the nebula may not decline exponentially with time; the favored process suggests it is highly super-exponential. That is, the nebula declines to essentially zero pressure in finite time, perhaps a mere 5 Ma. Accretion rates based on W isotopes are agnostic on the earliest stage of the Earth's growth but are usually modeled with longer timescales (**Figure 14**). However, noble gas data do provide evidence of

solar components within the Earth that could have been acquired directly from the solar nebula (Porcelli et al., 2001).

So long as the nebula persisted, therefore, the episodic magma ocean at the Earth's surface would have reacted with hydrogen gas, raising the Fe/FeO ratio of the bulk Earth and increasing the mass fraction of its metallic core. After about 10^7 years, the nebula dissipated and the oxidation state of the mantle would have been free to increase, particularly if late-arriving planetesimals had higher FeO/Fe ratio than the early ones (O'Neill, 1991a,b). This mechanism, impossible to quantify, would have generated a gradual increase in FeO/Fe ratio of the Earth, but would not have produced any of the Fe^{3+} present in peridotite and MORB (Wood, et al., 1990). The latter can be explained if the Earth self-oxidized through crystallization of magnesium silicate perovskite (Frost, et al., 2004; Wade and Wood, 2005).

At depths deeper than 660 km in the present-day Earth, the stable phases are $(Mg,Fe)SiO_3$ (magnesium silicate perovskite) (79% by volume; Wood, 2000), (Mg,Fe)O (magnesiowüstite) (16%), and $CaSiO_3$ perovskite (5%). An important property of magnesium silicate perovskite is that it accommodates the 5% Al_2O_3 that it dissolves in peridotite compositions by a coupled substitution with Fe^{3+} (McCammon, 1997; Wood and Rubie, 1996) as follows:

$$Mg^{2+}Si^{4+} \leftrightarrow Fe^{3+}Al^{3+}$$

It has recently been discovered that this substitution mechanism is so stable that it forces ferrous iron to disproportionate to ferric iron plus iron metal (Frost et al., 2004)

$$3Fe^{2+} \leftrightarrow 2Fe^{3+} + Fe^0$$

or, in terms of oxide components of the lower mantle,

$$3FeO + Al_2O_3 = 2FeAlO_3 + Fe^0$$

This means that, as perovskite began to crystallize in the extensively molten Earth, it dissolved ferric iron as $FeAlO_3$ component and produced Fe metal. Since the perovskite is stable throughout virtually the entire lower mantle, this process took place during most of accretion and core formation on Earth. The implications are that metal sinking through the lower mantle to the core would inevitably have dissolved some of this internally produced metallic iron, resulting in a perovskitic layer that was relatively oxidized. Given the gravitational instability of any metal layer (**Figure 16**), accretional energies, and heat loss, it is inevitable that the depth of the magma ocean fluctuated continuously, thereby generating fronts of dissolution and precipitation at the lower boundary. Perovskite dissolution and reprecipitation therefore released ferric iron to the magma ocean during dissolution and produced more during precipitation. This raised the oxygen fugacity of the mantle. Any later, droplets of metal falling through the magma ocean would have reacted with ferric iron, producing more Fe^{2+} that would then have dissolved in the silicate melt. Hence, the content of oxidized iron of the mantle (magma ocean) increased naturally as a consequence of perovskite crystallization. In the very final stages of the Earth's accretion, this mechanism may have caused sufficient oxidation to halt metal segregation, setting the stage for the 'late veneer' of chondritic or similar material to add the highly siderophile elements exclusively to the mantle. Note that the oxidation-state change required by partitioning data is a simple consequence of the size of the Earth. Any planet in which magnesium silicate perovskite is an important crystallizing phase should undergo the same process. It could explain why terrestrial basalts are more oxidized than those of Mars (Herd, et al., 2002), a planet in which perovskite can only crystallize at depths close to the core–mantle boundary. Because of its small size, Mars cannot have undergone the same period of extensive perovskite crystallization and self-oxidation during accretion as did Earth.

9.02.17 Isotopic Evidence for Impact-Induced Losses During the Earth's Accretion

Geochemistry provides three kinds of information on the changes that may have affected a planet during growth. The first is chemical compositions that imply some process of modification to primordial materials such as chondrites. The second is mass-dependent stable isotopic fractionation, which can leave a marked imprint of processes that affected the Earth. The third is isotopic compositions that leave a record of time-integrated parent/daughter elemental ratios that are distinct from those found in the Earth today.

The most dramatic example of mass-dependent fractionation is provided by Xe isotopes. Heavy Xe isotopes are affected by early decay of formerly live ^{129}I ($T_{½} = 16$ Ma) and ^{244}Pu ($T_{½} = 82$ Ma) (**Table 1**). The lighter isotopes with a mass of 124 to 128 are not affected by decay but are markedly fractionated in a mass-dependent fashion to heavier overall proportions relative to meteoritic and solar values (**Figure 10**). Originally, it was thought that this provided key evidence that the noble gases were more abundant in the early Earth and that a major fraction has been lost, probably by shock-induced blow off of a large protoatmosphere by hydrodynamic escape (Hunten et al., 1987; Walker, 1986). However, it is now clear that this effect is also found in the Earth's interior (Caffee et al., 1999; Holland et al., 2009) and a remarkably similar composition is also found in the martian atmosphere (Halliday, 2013) (**Figure 10**) despite a very different accretion history. Furthermore, neon (Ne), a much lighter and more readily fractionated element, displays no such sign of fractionation; it resembles a simple mixture of solar and chondritic components (**Figure 9**). Therefore, other models need to be considered. It would be helpful to know the Xe isotopic composition of the Venusian atmosphere because the martian and terrestrial composition may be a particular feature of terrestrial planets that relates to a specific kind of accretion/loss process. Although a loss process has not yet been clearly identified, late-stage parent/daughter fractionation has clearly affected the Earth's noble gases. Put simply, the small difference between the isotopic composition of xenon in Earth and its initial starting composition is consistent with lower parent/daughter ratios than are found, thereby implying a major loss of Xe given that late additions of the parents would make no sense. Live ^{129}I and ^{244}Pu must have been relatively low in abundance at the time of Xe loss because there is a shortage (by two orders of magnitude) of the decay products of ^{129}I and ^{244}Pu. Given the known abundances of I and refractory lithophile elements (of which Pu is an extinct example), the radiogenic and

fissionogenic Xe atoms that should have formed in the first 100 Ma say are largely missing (Porcelli and Pepin, 2000; Pepin and Porcelli, 2006). Therefore, there must have been a major fractionation in parent/daughter ratio at a relatively late stage and large amounts of loss. Early estimates for the age of this loss are 50–80 Ma (Ozima and Podosek, 1999; Porcelli and Pepin, 2000). However, longer timescales are required to satisfy the ^{136}Xe constraint and provide evidence that the Earth's Xe was lost over the first few hundred million years. Pepin and Porcelli (2006) had proposed two episodes of atmospheric blow off related to giant impacts. It is also possible that other processes operated in the Earth's atmosphere that may have caused loss, the most likely involving ionization of Xe (Halliday, 2013; Zahnle et al., 2006). Pujol et al. (2011) presented evidence of progressive loss of atmospheric Xe in the Earth's early history.

Supporting evidence for such losses comes from the composition of the major volatiles H, C, and N (Dasgupta and Hirschmann, 2010; Halliday, 2013). It has long been considered likely that major volatiles were added after the Moon-forming giant impact, and a number of papers (Albarède, 2009; Morbidelli et al., 2000) have made a connection to a late veneer that is thought to have been responsible for the abundances of highly siderophile elements in the Earth's mantle (Chou, 1978). However, the relative abundances of H, C, and N bear no resemblance to those of chondrites or comets (Halliday, 2013) (**Figure 18**). Therefore, some fractionating process must be responsible for their present concentrations. The volatiles have been lost either to space or to the Earth's core (Dasgupta and Hirschmann, 2010; Halliday, 2013). Core formation would be unlikely to leave the silicate Earth more depleted in nitrogen than in carbon (**Figure 19**). The most likely scenario therefore is that N, C, and H were differentially lost to space, possibly as a result of ionization along with Xe by solar EUV irradiation (**Figure 20**) (Halliday, 2013).

In addition to losses of elements like the noble gases and major volatiles, the question has arisen as to what extent the depletions in moderately volatile elements also reflect accretion phenomena. Evidence for loss of moderately volatile elements during accretion was hard to reconcile with the uniform K isotopic compositions recorded by inner solar system objects (Humayun and Clayton, 1995). However, the precision on these K isotopic measurements was not at the kind of level that would resolve small evaporation effects. Furthermore, if K were atmophile because it was degassed from the mantle into a hot atmosphere nebular atmosphere that was blown off with Xe (O'Neill, 1991a,b), the losses would not necessarily record a major isotopic fractionation (Halliday, 2004). Pritchard and Stevenson (2000) showed that in any case, it is possible to have large volatile loss without large isotopic effects. They present an explicit calculation showing how the K-41 isotopic effect can be small even when the volatilization is large. The main point is that it is evaporation into a dense atmosphere not into vacuum.

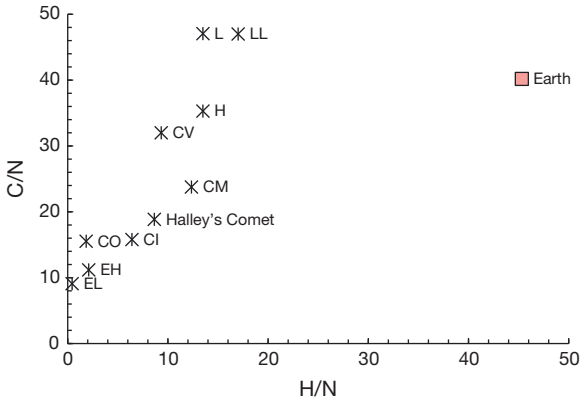


Figure 18 Weight ratios of C/N versus H/N in different classes of chondrites compared to the composition of the BSE. No combination of chondrites yields the composition of the Earth's major volatiles. Therefore, most H_2O (70%) predates the late veneer. Reproduced from Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.

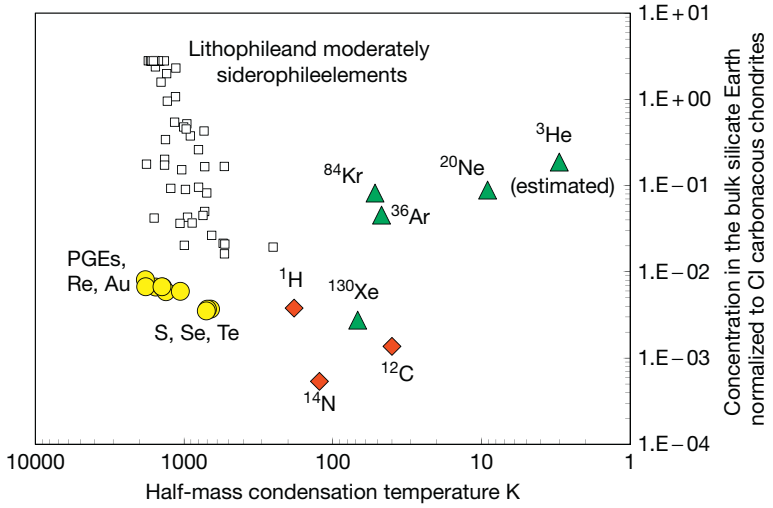


Figure 19 Concentrations of all elements in the BSE by weight normalized to CI carbonaceous chondrites, plotted against half-mass condensation temperatures given by Lodders (2003). Note the high concentrations of noble gases other than xenon and the strong depletion in carbon and nitrogen – more so than any other element in the BSE. Reproduced from Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.

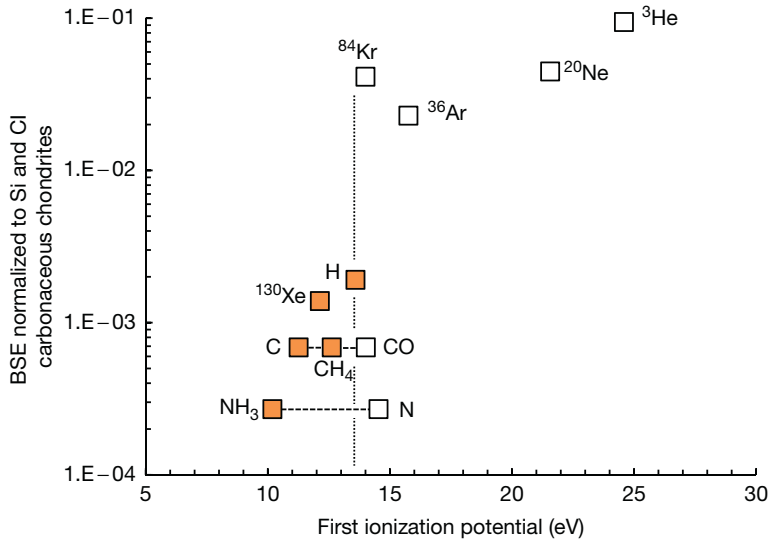


Figure 20 Concentrations of noble gases, hydrogen, carbon, and nitrogen in the BSE, normalized to silicon and CI and compared to the EUV first ionization potential. Differing ionization potentials are shown for differing molecular species of carbon and nitrogen, which would be a function of temperature (Lodders, 2003). Reproduced from Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.

It has been claimed that supporting evidence for accretionary losses of volatile elements is found in oxygen isotopic compositions (**Figure 3**). Oxygen isotopes provide a monitor of the degree of mixing of dust and gas in the disk (Clayton and Mayeda, 1975; Wiechert et al., 2001, 2004). The oxygen isotopic composition varies among different classes of meteorites and is therefore assumed to be highly heterogeneous in general among inner solar system objects (Clayton, 1986, 1993; Clayton et al., 1973). This is not a normal mass-dependent fractionation so much as a mass-independent effect, the origin of which is thought to be a self-shielding irradiation effect from the early Sun (Clayton, 2001). Regardless of the reason behind the heterogeneity, the oxygen isotopic composition provides a monitor of the origins of the material in a planetary object. The provenance of material in an object the size of the Earth may have been quite broad (Wetherill, 1994) although more recent accretion models define a more narrow feeding zone. As such, oxygen isotopes merely provide some average.

The Earth is distinct in oxygen isotopes from all classes of chondrites except enstatite chondrites, which may have formed in the more reducing inner regions of the solar system. Similarly, the provenance appears to be distinct from that of the material forming asteroid 4 Vesta or Mars, as judged from studies of meteorites thought to be derived from these objects (Clayton, 1986, 1993; Wiechert et al., 2004). The Moon, however, has an oxygen isotopic composition that is identical to that of the Earth (Clayton and Mayeda, 1975), demonstrated to persist to extremely high precision (Wiechert et al., 2001) (**Figure 3**). This has long provided a powerful line of evidence that the Moon formed from the Earth (Ringwood, 1989a,b; Wänke and Dreibus, 1986) or that the Earth and Moon formed from the same material (Wiechert et al., 2001), even though these two objects have demonstrably different chemical composition.

In simulations of the giant impact (Cameron and Benz, 1991), roughly 80% of the Moon forms from the debris

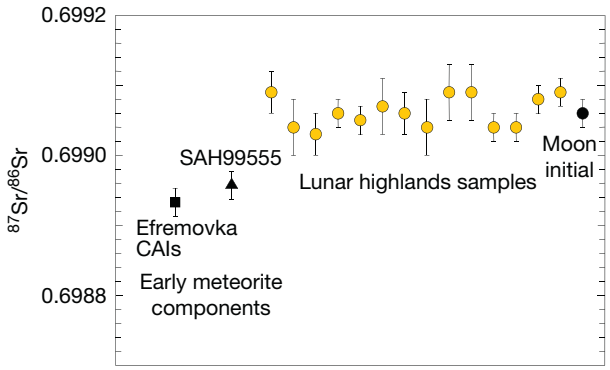


Figure 21 Initial Sr isotopic composition of early lunar highland rocks relative to other early solar system objects. From compilation in Halliday and Porcelli (2001). APB, angrite parent body; CEPB, cumulate eucrite parent body; BSSI, bulk solar system initial.

derived from Theia (Canup and Asphaug, 2001). If this is correct, the oxygen isotopic evidence provides evidence that Theia and the Earth had similar average provenance. Yet the Moon is far more depleted in volatile elements than the Earth (Taylor and Norman, 1990) (**Figure 4**). A possible explanation for this is that there were major reductions in volatile constituents during the Moon-forming giant impact (O'Neill, 1991b). Support for this is found in the initial $^{87}Sr/^{86}Sr$ of early lunar rocks from the highlands (Halliday and Porcelli, 2001). These yield an initial $^{87}Sr/^{86}Sr$ for the Moon that is clearly resolvable from the initial composition of the solar system (**Figure 21**), consistent with radiogenic growth in an environment with relatively high Rb/Sr.

It has been argued that it is very unlikely that the Earth and Theia had similar provenance. Pahlevan and Stevenson (2007) and Pahlevan et al. (2011) proposed that the identical oxygen isotopic compositions of the Earth and Moon more likely

represent equilibration of atoms between the hot proto-lunar disk and the Earth's magma ocean in the aftermath of the giant impact. As explained earlier in the text, it is now known that Si (Armytage et al., 2012), Ti (Leya et al., 2008; Zhang et al., 2012), Cr (Lugmair and Shukolyukov, 1998), and W (Touboul et al., 2007) are identical between the Earth and Moon but variable within the broader solar system. Therefore, the case for the Moon's atoms being largely derived from the Earth whether by direct impact or by equilibration is overwhelming.

Impact-driven chemical changes in the Earth are well supported by evidence that the planet Mercury lost a major fraction of its silicate by impact erosion (Benz et al., 1987). The Earth also has a Fe/Mg ratio of at least 2.11, which is 10% higher than the value of 1.90 for CI chondrites (Palme, 2000). A possible explanation for this is impact erosion of the outer portions of the Earth (Halliday et al., 2001; O'Neill and Palme, 2008). Impact erosion is likely to generate a Sm/Nd that is higher than chondritic through loss of low Sm/Nd crust, and with new high-precision isotopic data for meteorites, evidence for a subtle but important difference between chondrites and the Earth has been growing (Boyet and Carlson, 2005). This has spawned new questions about whether the Earth is simply nonchondritic (Drake and Righter, 2002) or whether there are hidden reservoirs inside the Earth that balance its composition (Boyet and Carlson, 2005) (see next section). The observed ^{142}Nd effect is explicable with just a 6% increase in Sm/Nd in the BSE relative to chondrites (Halliday, 2006).

O'Neill and Palme (2008) made the case that there have been selective losses of all incompatible elements from the outer portions of the Earth as a result of such impact erosion. If this is the case, the calculated budgets of refractory incompatible elements like Th and U have been overestimated and could be 40–50% lower. Potassium budgets are based on the K/U ratio of the mantle so would also be correspondingly reduced. Radiogenic heat production would be much less than currently accepted if, as argued, there has been impact erosion. The K budget affects the calculated budget for radiogenic ^{40}Ar in the Earth, which in turn provides a quantification of the ^{36}Ar budget so the small amount of primordial volatiles is even smaller if there has been impact erosion. The budgets for major volatiles are in turn determined by being ratioed to those of Ar isotopes (Marty, 2012) so the amount of water in the mantle, for example, could be as different as 1 or 9 ocean masses (Halliday, 2013) depending on impact erosion.

The mismatch between U–Pb chronometry and Hf–W chronometry may also be related to late volatile losses. The rates of accretion of the Earth based on the model of **Figure 14** and the various estimates of the BSE Pb isotopic composition (**Figure 11**) are shown in **Figure 15**. Also shown are the results for W. It is clear that the kinds of timescales involved are on the order of 10^7–10^8 years, providing strong vindication of the Safronov–Wetherill style of protracted accretion via major impacts (Halliday, 2014). It is also immediately apparent however that a discrepancy exists between the W results and every estimate based on Pb isotopic data (Halliday, 2004). Either all of these Pb isotopic estimates are incorrect or there is some basic difference in the way these clocks operate. There are a number of parameters that play into these models that have a significant level of uncertainty associated.

When the ^{182}Hf–^{182}W and $^{235/238}$U–$^{207/206}$Pb chronometers are compared using the same style of model, a variable but unidirectional offset is found between the timing that is based on W and that based on any of the 11 estimates of the Pb isotopic composition of the silicate Earth (**Figure 11**). For example, the two-stage ^{182}Hf–^{182}W model age of the Earth is 30 Ma (Yin et al., 2002), whereas the same model applied to $^{235/238}$U–$^{207/206}$Pb yields ages ranging between >40 and >100 Ma (**Figure 11**). It is of course not unlikely that all the Pb isotopic estimates are wrong. Given the difficulties in defining meaningful average Pb isotopic ratios for the silicate Earth, this is certainly a possibility. In fact, some more recent estimates of Pb isotopic composition of the BSE have been made based on the results of W isotopes so there is circularity involved. Assuming, however, that some of the Pb isotopic estimates are close to correct, one can consider what kind of processes might result in a difference between W and Pb ages of the core.

One explanation (Wood and Halliday, 2005, 2010) is that the transfer of W and Pb to the core changed, but not together, during the accretion history of the Earth. Tungsten is moderately siderophile but not chalcophile, whereas Pb is thought to tend toward the opposite. Following the cooling and oxidation of the Earth after the giant impact, sulfide is likely to have formed. Removal of this sulfide to the core may have been responsible for a late-stage increase in U/Pb that defines the Pb isotopic compositions observed. This being the case, the Pb isotopic composition of the BSE provides information on the timescales over which the Earth cooled and core segregated following the giant impact. The timescales inferred depend on which Pb isotopic estimate is employed (**Figures 11** and **15**). However, they are all of the order of 100 Ma, which is close to the age of the giant impact.

There is another possible explanation for the apparent discrepancy between accretion rates determined from ^{182}Hf–^{182}W to $^{235/238}$U–$^{207/206}$Pb. Lead is moderately volatile, and as such, the younger ages could reflect loss of Pb from the Earth during accretion. Loss of such a heavy element from the Earth is dynamically difficult. However, if the surface of the Earth was very hot with a magma ocean, Pb might have been episodically atmophile rather than lithophile. The atmosphere may then have been blown off at a late stage (Walker, 1986). The depletion of the terrestrial planets in moderately volatile elements has long been considered to be a result of incomplete condensation because it was believed that the inner solar system formed from a hot solar nebular gas. Perhaps some of this depletion is, in fact, a late-stage phenomenon. Support for this model was found in results for the new ^{205}Pb–^{205}Tl chronometer (Nielsen et al., 2006) that are hard to explain without major losses of Tl from the Earth at a late stage.

With the advent of MC-ICPMS (Halliday et al., 1998), it has become possible to measure small mass-dependent isotopic differences between early solar system objects that may have been induced by the processes of accretion and evaporation of silicate and metal. Of particular interest are the less volatile but not highly refractory elements that would not have formed part of the Earth's protoatmosphere. Such elements include Li, Mg, Si, V, Cr, Fe, and Ni (Lodders, 2003). Precise stable isotopic compositions for mantle-derived samples from the Earth and meteorite groups have now been published for Li (Magna et al., 2006; Seitz et al., 2006), Mg (Pogge von Strandmann et al., 2011; Wiechert and Halliday, 2007), and Fe (Craddock et al., 2013; Poitrasson et al., 2004; Weyer et al., 2005; Williams

et al., 2004, 2006). The Fe isotopic data for 'normal' basalts from the Earth and Moon are slightly heavy (roughly 40 ppm per amu) relative to data for Mars and Vesta (Poitrasson et al., 2004; Weyer et al., 2005). This would be consistent with the loss of small amount of evaporated Fe during the giant impact. However, Williams et al. (2004) had shown that isotopic fractionations may also result from oxidation effects. Basalt melting may have generated different isotopic fractionations on the Earth, Moon, and other objects (Craddock et al., 2013; Schoenberg and von Blanckenburg, 2006). It has also been argued that the silicate Earth, being larger than Mars or Vesta, may have generated a heavy Fe isotopic composition during partitioning ferric iron into perovskite in equilibrium with metal that was transferred to the core, as discussed in the preceding text (Williams et al., 2006, 2012).

Silicon (Georg et al., 2007) and vanadium (Nielsen et al., 2014; Prytulak et al., 2013) display clear isotopic differences between the silicate Earth and meteorites although there have been counter claims for Si (Chakrabarti and Jacobsen, 2010; Fitoussi et al., 2009) that have not been replicated with more detailed work (Armytage et al., 2011; Savage et al., 2010). In the case of Si, the small (0.1‰ per amu) effect appears to relate to core formation and an isotopic fractionation between Si in metal and silicate (**Figure 22**). This is supported by experimental studies. In the case of V, the effect is very large (1‰ per amu) and no evidence for such a metal–silicate fractionation has yet been found (Nielsen et al., 2014). The explanation may relate to irradiation in the early circumstellar disk.

Evidence that the Moon does not carry a heavy isotopic composition as a result of boiling during the giant impact is found in the Li and Mg isotopic compositions (Magna et al., 2006; Seitz et al., 2006; Wiechert and Halliday, 2007), which provide no hint of a difference between any of the inner solar system-differentiated planetary objects.

9.02.18 Hidden Reservoirs and the Composition of the Earth

In all of the foregoing discussion, it is assumed that the composition of the BSE as deduced from basalts is representative of the entire silicate Earth. In fact, it may be different from chondrites or our best estimates of its true composition if there are *hidden reservoirs*. It has been argued that this is indeed the case, on the basis of oxygen isotopes. The oxygen isotopic heterogeneity of the solar system is not only of importance for defining meteorite parent bodies. It provides a fingerprint for the material that formed the Earth. The similarity between the Earth and Moon in terms of oxygen isotopes provides powerful evidence for a similar average provenance or equilibration. It is perhaps significant that the only chondrites with an oxygen isotopic composition like that of Earth are the enstatite chondrites. This has led to a variety of Earth models based on a total composition that is highly reduced and similar to that of the enstatite chondrites (e.g., Javoy, 1999). Because the mantle and crust of the Earth sampled by magmas and xenoliths are inconsistent with this composition, the 'enstatite chondrite Earth' has to have hidden reservoirs (customarily located in the lower mantle and hence not sampled) that are compositionally distinct from anything we can readily measure.

Hidden reservoir models are of little use scientifically because they are untestable. Further evidence of a link between the Earth and enstatite chondrites has, however, come from nitrogen (Marty et al., 2003) and chromium (Shukolyukov and Lugmair, 2004) isotopic measurements, adding support to the 'hidden reservoir' concept. Recently, a hidden reservoir of different composition has been proposed by Tolstikhin and Hofmann (2005) to explain noble gas isotope systematics. More compelling is the evidence that the BSE as sampled is slightly different from chondrites in terms of ^{142}Nd atomic abundance, possibly reflecting a different ^{146}Sm/^{142}Nd in the early solar system (Boyet and Carlson, 2005). One explanation given is that the Earth contains a hidden primordial reservoir with complementary lower ^{146}Sm/^{142}Nd balancing the higher ^{146}Sm/^{142}Nd of the silicate Earth as sampled by basalts and continents. The apparent model age of this reservoir is 30 Ma after the start of the solar system but this would have to predate the Moon-forming giant impact, which seems unlikely.

Melting as a result of the giant impact, followed by settling of majoritic garnet or silicate perovskites in the magma ocean, has been proposed as a mechanism for forming the primitive upper mantle with high, nonchondritic Mg/Si ratio (e.g., Herzberg et al., 1988; Ohtani et al., 1986; Ringwood, 1979)

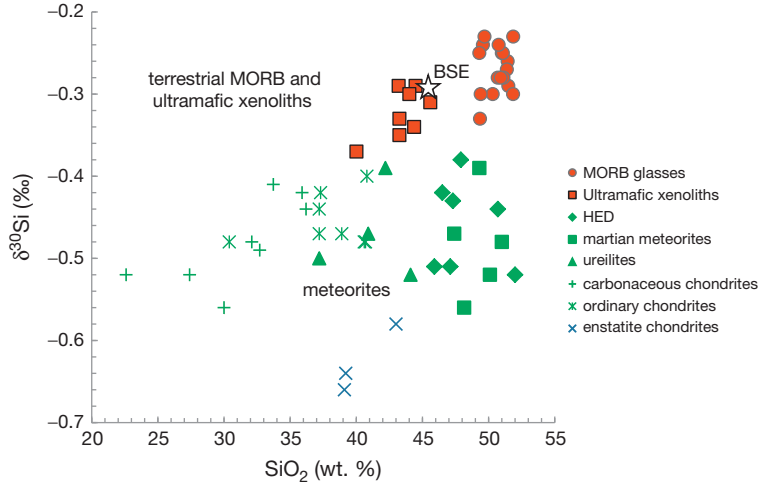


Figure 22 Silicon isotopic compositions of meteorites (Armytage et al., 2011) compared to average data defining the composition of the BSE (Savage et al., 2010).

and a complementary 'hidden' lower mantle reservoir. If substantial fractional crystallization of the magma ocean occurred, however, subsequent remixing must have removed most of the material segregated to the lower mantle because of the observation that the ratios of the refractory lithophile elements in the upper mantle are chondritic. The question is, however, could a small amount of undetected fractionation have shifted the Sm/Nd ratio of the silicate Earth sufficiently to generate the observed ^{142}Nd anomaly? Crystal–melt partition coefficients for majorite and for $CaSiO_3$ and $(Mg,Fe)SiO_3$ perovskites have been measured by Corgne and Wood (2004), Liebske et al. (2005), and Walter et al. (2004) and applied to calculate the maximum volumes of 'hidden' majorite or perovskite that could be present in the lower mantle, given the compositional constraints derived from the upper mantle. In the case of majorite, the maximum allowable hidden reservoir is 14% of the mantle volume (Corgne and Wood, 2004) while the maximum extent of combined calcium and magnesium silicate perovskite fractionation is about 10% (Liebske et al., 2005; Walter et al., 2004). In neither case is the 'hidden' amount sufficient to explain the nonchondritic Mg/Si ratio of the observable BSE (**Figure 23**).

Interestingly, the hidden reservoir would in both cases have a higher Sm/Nd ratio than the BSE, which is the wrong direction to explain the ^{142}Nd anomaly of the BSE. This leads, however, to the possibility that dense high-pressure melts (complementary to solid majorite and perovskite) could remain isolated at great depths and generate hidden reservoirs (e.g., Suzuki and Ohtani, 2003, and references therein). Recent molecular dynamics simulations in the $MgSiO_3$ system (Stixrude and Kirko, 2005) suggest that the density contrast between melt and crystals decreases with increasing pressure, such that a density crossover near the base of the mantle could occur. Hofmann et al. (2005) had proposed that downward migration of such dense melts formed in the deep mantle in equilibrium with Ca perovskite could have created a deep 'missing' reservoir unradiogenic in Pb (with a U/Pb ratio lower than that of PUM) as required to balance the Pb composition of the bulk MORB and OIB source reservoirs. It would have high Nd/Sm ratio and could hence be the missing subchondritic ^{142}Nd/^{144}Nd reservoir (Boyet and Carlson, 2005).

There are, however, problems preserving any distinct mantle reservoir from this early in the Earth's history. After the Moon formed, there would have been tidal interactions with the Earth that would have caused strong heating of the lower mantle in particular (Zahnle et al., 2006). Mantle overturn and melting should have been especially vigorous eliminating any early heterogeneity. Therefore, explanations for the Nd isotopic based on hidden reservoirs are at present very tentative.

9.02.19 Concluding Overview

1. The solar system formed because the angular momentum of the collapsing portion of a molecular cloud was transferred to a swirling disk of gas and debris left over from the proto-Sun. The disk experienced strong heating from the T-Tauri stage of irradiation and from frictional heating as material was being rapidly swept in toward the hot new Sun. This led to strong depletions in moderately volatile elements close in to the Sun.
2. The Earth was built from differentiated planetary embryos that formed from this volatile-depleted dust in the inner solar system. Such embryos and their metallic cores formed extremely early. The latest ^{82}Hf–^{182}W chronometry provides evidence of timescales of <2 Ma for iron meteorite parent bodies. Mars may also have formed this quickly. These bodies almost certainly melted and differentiated as a result of the decay of short-lived nuclides in particular ^{26}Al and ^{60}Fe.
3. Chondrites are the most abundant form of meteorite that strikes the Earth. They have relative concentrations of nonvolatile elements that are broadly similar to those found in the Sun. They represent a convenient reference for the bulk composition of the Earth. However, though primitive, they are not early. They must have formed at least 1–3 Ma after the start of the solar system because the

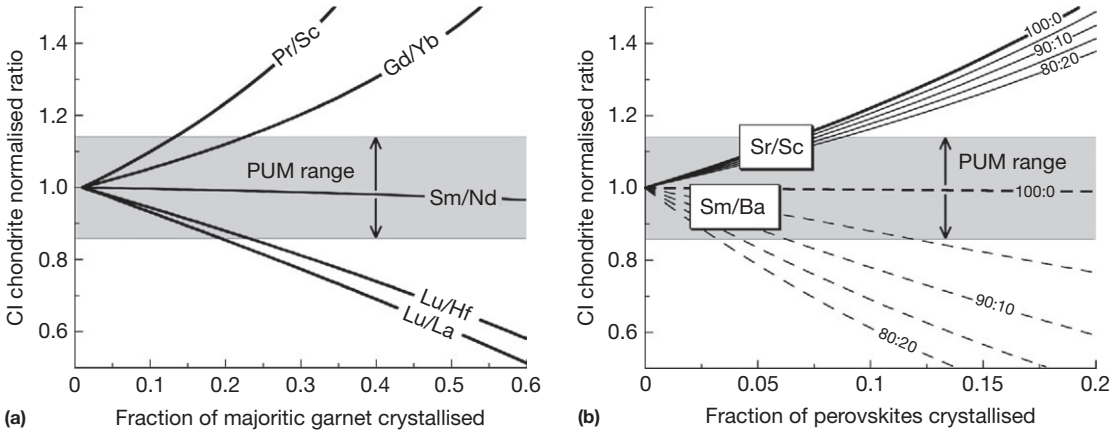


Figure 23 The effects of perovskite fractionation on Sm/Nd and Lu/Hf ratios.

chondrules they contain provide evidence of an epoch of formation in the dusty disk or solar nebula at this time.

4. The Earth accreted via a stochastic series of planetary collisions over tens of millions of years. The final major accretion stage was the Moon-forming giant impact currently estimated at between 70 and 110 Ma after the start of the solar system, when the last roughly 10% of the Earth was added.
5. The chemical composition of the solar system is defined using spectroscopic measurements of the Sun combined with laboratory measurements of chondritic meteorites. All of the stable elements and isotopes are present within our solar system and their relative abundance provides powerful evidence in support of the theories of nuclear stability in the stars and of stellar nucleosynthesis.
6. The density of the Earth combined with estimates of its core size leads to the conclusion that it must be approximately chondritic in its relative concentrations of refractory elements. Similarly, the chemical composition of the BSE, which is well constrained from measurements of mantle peridotites and basalts, exhibits ratios of refractory lithophile elements that are the same as in chondritic meteorites, the Moon, and the silicate portion of Mars.
7. The BSE is depleted in volatile elements relative to chondrites, and it has Mg/Si slightly greater and Mg/refractory lithophile elements slightly lower than chondritic. Depletions of the BSE in refractory siderophile elements are due to their sequestration in the core.
8. By using the chondritic reference model for refractory elements in the Earth, we can calculate the partitioning of siderophile elements between the core and the mantle. The results may then be used in conjunction with experimental liquid metal–liquid silicate partition coefficients to estimate the conditions of core segregation. This exercise requires the assumption that almost all of the core formed in equilibrium with the mantle. Partition coefficients depend on pressure and temperature, and the depletions of the BSE in siderophile elements can only be matched by very high pressures and temperatures.
9. Results imply segregation of the core at the base of a deep magma ocean that extended to 30–40% of the depth of the mantle in the growing earth. The bottom of the magma ocean was by definition below the peridotite liquidus temperature and the constraints this imposes leads to the requirement that the Earth became more oxidized during accretion. The BSE composition can only be matched if the oxidation state of the Earth changed as it grew. The resultant core contains about 4% Si, thus indicating that it contains substantial Si and is a major reason for the superchondritic Mg/Si of the BSE.
10. As the Earth accreted, it would, when it reached the current-day size of Mars, have begun to crystallize silicate perovskites close to the core–mantle boundary. The strong affinity of $(Mg,Fe)SiO_3$ perovskite for ferric iron means that crystallization of this phase forces disproportionation of Fe^{2+} into Fe^{3+} plus metal. This is one mechanism for oxidizing the Earth during accretion since progressive growth through impact would have led to the release of Fe^{3+} to the magma ocean and of metal to the core. This mechanism also could explain the observation that the mantle of the Earth is at a higher oxygen fugacity than the mantles of Mars and the Moon.
11. The Moon-forming giant impact did not result in any significant losses of slightly volatile elements such as Li, Mg, and Fe.
12. The similarity between the O, Si, Ti, Cr, and W isotopic compositions of the Earth and Moon could have resulted from isotopic equilibration between the materials accreting to form the Moon with the proto-Earth. An alternative is that radically different giant impact scenarios need to be considered.
13. Major losses of volatiles affected Earth during its accretion, including losses of early-formed atmospheres. Such losses are apparent from Xe isotopic compositions and the strongly nonchondritic proportions of H/C/N in the silicate Earth ratios.
14. The late U–Pb age for the Earth is consistent with a giant impact at 70–110 Ma. The difference between Hf–W chronology and U–Pb chronology may reflect changes in partitioning during the final stages of core formation, or W isotopes reflect the effects of disequilibrium during accretion by core–core merging.
15. There is evidence that the Earth may be nonchondritic because of impact erosion of its outer portions during accretion. If this is the case, the budgets for incompatible elements including all heat-producing elements U, Th, and K have been overestimated, potentially by almost a factor of 2.
16. Budgets for hidden volatiles are closely related to those of U and K. Impact erosion models would lead to calculated mantle–water budgets of about 1 ocean mass, whereas standard chondritic models yield mantle–water budgets of nine ocean masses.
17. After the giant impact, crystallization of the Earth would have led to the production of majoritic garnet in the transition zone and of $(Mg,Fe)SiO_3$ and $CaSiO_3$ perovskites in the lower mantle. Crystal–liquid partition coefficients for these phases can be used to estimate the extent of any long-lived 'hidden' reservoir produced in the lower mantle after mantle crystallization. Results indicate a maximum extent of majorite crystallization of 14% of mantle volume and of perovskite fractionation of 10% could be hidden without perturbing the chondritic refractory lithophile element ratios of the BSE.
18. A 'hidden' reservoir with superchondritic Nd/Sm required to balance a putative chondritic BSE could not be perovskite or majorite since these phases all prefer Sm over Nd. One possibility would be isolation and freezing of dense high-pressure melts (complementary to perovskite) in the lower mantle. Such a low-volume reservoir would have appropriate Nd/Sm to generate superchondritic $^{142}Nd/^{144}Nd$ in the remaining BSE. However, the preservation of such a reservoir is hard to reconcile with the strong tidal melting anticipated for the deep Earth following Moon formation. Furthermore, the ^{142}Nd effects can be explained by small isotopic or elemental (Sm/Nd) differences between the Earth and chondrites, as exemplified by other lines of evidence.

Acknowledgments

BJW and ANH acknowledge the support of STFC and ERC for their research programs at Oxford.

References

Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.

Agnor CB and Ward WR (2002) Damping of terrestrial-planet eccentricities by density–wave interactions with a remnant gas disk. *Astrophysical Journal* 567: 579–586.

Albarède F (2009) Volatile accretion history of the terrestrial planets and dynamic implications. *Nature* 461: 1227–1233.

Allègre CJ and Lewin E (1989) Chemical structure and history of the Earth: Evidence from global non-linear inversion of isotopic data in a three box model. *Earth and Planetary Science Letters* 96: 61–88.

Allègre CJ, Lewin E, and Dupré B (1988) A coherent crust–mantle model for the uranium–thorium–lead isotopic system. *Chemical Geology* 70: 211–234.

Allègre CJ, Manhès G, and Göpel C (1995a) The age of the Earth. *Geochimica et Cosmochimica Acta* 59: 1445–1456.

Allègre CJ, Poirier J-P, Humler E, and Hofmann AW (1995b) The chemical composition of the Earth. *Earth and Planetary Science Letters* 134: 515–526.

Amelin Y, Krot AN, Hutcheon ID, and Ulyanov AA (2002) Lead isotopic ages of chondrules and calcium–aluminum-rich inclusions. *Science* 297: 1678–1683.

Arculus RJ and Delano JW (1981) Siderophile element abundances in the upper mantle: Evidence for a sulfide signature and equilibrium with the core. *Geochimica et Cosmochimica Acta* 45: 1331–1344.

Armytage RMG, Georg RB, Savage PS, Williams HM, and Halliday AN (2011) Silicon isotopes in meteorites and planetary core formation. *Geochimica et Cosmochimica Acta* 75: 3662–3676.

Armytage RMG, Georg RB, Williams HM, and Halliday AN (2012) Silicon isotopes in lunar rocks: Implications for the Moon's formation and the early history of the Earth. *Geochimica et Cosmochimica Acta* 77: 504–514.

Arnould M, Paulus G, and Meynet G (1997) Short-lived radionuclide production by non-exploding Wolf–Rayet stars. *Astronomy and Astrophysics* 321: 452–464.

Benz W, Cameron AGW, and Slattery WL (1987) Collisional stripping of Mercury's mantle. *Icarus* 74: 516–528.

Bizzarro M, Baker JA, and Haack H (2004) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* 431: 275–278.

Bizzarro M, Baker JA, and Haack H (2005) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions – Correction. *Nature* 435: 1280.

Blum J (2000) Laboratory experiments on preplanetary dust aggregation. *Space Science Reviews*, pp. 265–278, Springer.

Borg LE, Connelly JN, Boyet M, and Carlson RW (2011) Chronologic evidence that the Moon is either young or did not have a global magma ocean. *Nature* 477: 70–72.

Boss AP (1990) 3D Solar nebula models: Implications for Earth origin. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 3–15. Oxford: Oxford University Press.

Boss AP (1997) Giant planet formation by gravitational instability. *Science* 276: 1836–1839.

Bottke WF, Nesvorný D, Grimm RE, Morbidelli A, and O'Brien DP (2006) Iron meteorites as remnants of planetesimals formed in the terrestrial planet region. *Nature* 439: 821–824.

Boyet M and Carlson RW (2005) ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.

Brearley AJ and Jones RH (1998) Chondritic meteorites. *Reviews in Mineralogy 36: Planetary Materials*, pp. 1-398. Washington, DC: Mineralogical Society of America, chapter 3.

Broecker WS (1998) *How to Build a Habitable Planet*. Columbia University, New York: Eldigio Press.

Burbidge GR, Burbidge EM, Fowler WA, and Hoyle F (1957) Synthesis of elements in stars. *Reviews of Modern Physics* 29: 547–650.

Busso M, Gallino R, and Wasserburg GJ (2003) Short-lived nuclei in the early Solar System: A low mass stellar source? *Publications of the Astronomical Society of Australia* 20: 356–370.

Caffee MW, Hudson GB, Velsko C, Huss GR, Alexander EC Jr., and Chivas AR (1999) Primordial noble gases from Earth's mantle: Identification of a primitive volatile component. *Science* 248: 2115–2118.

Cameron AGW and Benz W (1991) Origin of the Moon and the single impact hypothesis IV. *Icarus* 92: 204–216.

Cameron AGW and Truran JW (1977) The supernova trigger for formation of the Solar System. *Icarus* 30: 447–461.

Canup RM and Agnor CB (2000) Accretion of the terrestrial planets and the Earth–Moon system. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.

Canup RM and Asphaug E (2001) Origin of the moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.

Carlson RW and Lugmair GW (1988) The age of ferroan anorthosite 60025: Oldest crust on a young Moon. *Earth and Planetary Science Letters* 90: 119–130.

Cassen P (1996) Models for the fractionation of moderately volatile elements in the solar nebula. *Meteoritics and Planetary Science* 31: 793–806.

Chabot NL and Agee CB (2003) Core formation in the Earth and Moon: New experimental constraints from V, Cr, and Mn. *Geochimica et Cosmochimica Acta* 67(11): 2077–2091.

Chakrabarti R and Jacobsen SB (2010) Silicon isotopes in the inner solar system: Implications for core formation, solar nebular processes and partial melting. *Earth and Planetary Science Letters* 74: 6921–6933.

Chambers JE (2004) Planetary accretion in the inner Solar System. *Earth and Planetary Science Letters* 223: 241–252.

Chou CL (1978) Fractionation of siderophile elements in the earth's upper mantle. *Proceedings of Lunar and Planetary Science Conference* 9: 219–230.

Ciesla FJ and Charnley SB (2006) The physics and chemistry of nebular evolution. In: Lauretta D, Leshin L, and MacSween H (eds.) *Meteorites and the Early Solar System II*, pp. 209–230. Tucson, AZ: University of Arizona Press.

Clayton RN (1986) High temperature isotope effects in the early Solar System. In: Valley JW, Taylor HP, and O'Neil R (eds.) *Stable Isotopes in High Temperature Geological Processes*, pp. 129–140. Washington, DC: Mineralogical Society of America.

Clayton RN (1993) Oxygen isotopes in meteorites. *Annual Review of Earth and Planetary Sciences* 21: 115–149.

Clayton RN (2001) Self-shielding in the solar nebula. *Nature* 415: 860–861.

Clayton RN, Grossman L, and Mayeda TK (1973) A component of primitive nuclear composition in carbonaceous meteorites. *Science* 182: 485–487.

Clayton RN and Mayeda TK (1975) Genetic relations between the moon and meteorites. *Proceedings of Lunar and Planetary Science Conference* IX: 1761–1769.

Corgne A and Wood BJ (2004) Trace element partitioning between majoritic garnet and silicate melt at 25 GPa. *Physics of the Earth and Planetary Interiors* 143–144: 407–419.

Craddock PR, Warren JM, and Dauphas N (2013) Abyssal peridotites reveal the near-chondritic Fe isotopic composition of the Earth. *Earth and Planetary Science Letters* 365: 63–76.

Cuzzi JN, Hogan RC, Paque JM, and Dobrovolskis AR (2001) Size-selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophysical Journal* 546: 496–508.

Dahl TW and Stevenson DJ (2010) Turbulent mixing of metal and silicate during planet accretion—And interpretation of the Hf–W chronometer. *Earth and Planetary Science Letters* 295: 177–186.

Dasgupta R and Hirschmann MM (2010) The deep carbon cycle and melting in Earth's interior. *Earth and Planetary Science Letters* 298: 1–13.

Dauphas N and Pourmand A (2011) Hf–W–Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* 473: 489–492.

Davies GF (1984) Geophysical and isotopic constraints on mantle convection: An interim synthesis. *Journal of Geophysical Research* 89: 6017–6040.

Desch SJ, Morris MA, Connolly HC Jr., and Boss AP (2010) A critical examination of the X-wind model for chondrule and calcium-rich, aluminum-rich inclusion formation and radionuclide production. *Astrophysical Journal* 725: 692.

Doe BR and Zartman RE (1979) In: Barnes HL (ed.) *Geochemistry of Hydrothermal Ore Deposits*, pp. 22–70. New York: Wiley.

Drake MJ and Righter K (2002) Determining the composition of the Earth. *Nature* 416: 39–44.

Feigelson ED, Broos P, Gaffney JA III, et al. (2002a) X-ray-emitting young stars in the Orion nebula. *Astrophysical Journal* 574: 258–292.

Feigelson ED, Garmire GP, and Pravdo SH (2002b) Magnetic flaring in the pre-main-sequence sun and implications for the early Solar System. *Astrophysical Journal* 572: 335–349.

Fitoussi C, Bourdon B, Kleine T, Oberli F, and Reynolds BC (2009) Si isotope systematics of meteorites and terrestrial peridotites: Implications for Mg/Si fractionation in the solar nebula and for Si in the Earth's core. *Earth and Planetary Science Letters* 287: 77–85.

Foley CN, Wadhwa M, Borg LE, Janney PE, Hines R, and Grove TL (2005) The early differentiation history of Mars from ^{182}W–^{142}Nd isotope systematics in the SNC meteorites. *Geochimica et Cosmochimica Acta* 69: 4557–4571.

Foster PN and Boss AP (1996) Triggering star formation with stellar ejecta. *Astrophysical Journal* 468: 784–796.

Foster PN and Boss AP (1997) Injection of radioactive nuclides from stellar source that triggered the collapse of the presolar nebula. *Astrophysical Journal* 489: 346–357.
Frost DJ, Liebske C, Langenhorst F, McCammon CA, Trønnes RG, and Rubie DC (2004) Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle. *Nature* 428: 409–412.
Galer SJG and Goldstein SL (1991) Depleted mantle Pb isotopic evolution using conformable ore leads. *Terra Abstracts (EUG VI)* 3: 485–486.
Galer SJG and Goldstein SL (1996) Influence of accretion on lead in the Earth. In: Basu AR and Hart SR (eds.) *Isotopic Studies of Crust–Mantle Evolution*, pp. 75–98. Washington, DC: AGU.
Gast PW (1960) Limitations on the composition of the upper mantle. *Journal of Geophysical Research* 65: 1287–1297.
Georg RB, Halliday AN, Schauble E, and Reynolds BC (2007) Silicon in the Earth's core. *Nature* 447: 1102–1106.
Gessmann CK and Rubie DC (1998) The effect of temperature on the partitioning of nickel, cobalt, manganese, chromium, and vanadium at 9 GPa and constraints on formation of the Earth's core. *Geochimica et Cosmochimica Acta* 62: 867–882.
Gessmann CK, Wood BJ, Rubie DC, and Kilburn MR (2001) Solubility of silicon in liquid metal at high pressure: Implications for the composition of the Earth's core. *Earth and Planetary Science Letters* 184: 367–376.
Goldreich P and Ward WR (1973) The formation of planetesimals. *Astrophysical Journal* 183: 1051–1060.
Gounelle M, Shu FH, Shang H, Glassgold AE, Rehm KE, and Lee T (2001) Extinct radioactivities and protosolar cosmic-rays: Self-shielding and light elements. *Astrophysical Journal* 548: 1051.
Halliday AN (2000) Terrestrial accretion rates and the origin of the Moon. *Earth and Planetary Science Letters* 176: 17–30.
Halliday AN (2004) Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* 427: 505–509.
Halliday AN (2006) The origin of the Earth – What's new? *Elements* 2: 205–210.
Halliday AN (2008) A young Moon-forming giant impact at 70 to 110 million years accompanied by late-stage mixing, core formation and degassing of the Earth. *Philosophical Transactions of the Royal Society of London, Series A* 366: 4163–4181.
Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.
Halliday AN (2014) The origin and earliest history of the Earth. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry,* 2nd edn, vol. 2, pp. 149–211. Oxford: Elsevier.
Halliday AN and Kleine T (2006) Meteorites and the timing, mechanisms and conditions of terrestrial planet accretion and early differentiation. In: Lauretta D, Leshin L, and MacSween H (eds.) *Meteorites and the Early Solar System II*, pp. 775–801. Tucson, AZ: University of Arizona Press.
Halliday AN, Lee DC, Christensen JN, et al. (1998) Applications of multiple collector ICPMS to cosmochemistry, geochemistry and paleoceanography (The 1997 Geochemical Society Presidential Address.). *Geochimica et Cosmochimica Acta* 62: 919–940.
Halliday AN, Lee D-C, Porcelli D, Wiechert U, Schönbächler M, and Rehkämper M (2001) The rates of accretion, core formation and volatile loss in the early Solar System. *Philosophical Transactions of the Royal Society* 359: 2111–2135.
Halliday AN and Porcelli D (2001) In search of lost planets – The paleocosmochemistry of the inner Solar System. *Earth and Planetary Science Letters* 192: 545–559.
Hanan BB and Tilton GR (1987) 60025: Relict of primitive lunar crust? *Earth and Planetary Science Letters* 84: 15–21.
Hanks TC and Anderson DL (1969) The early thermal history of the earth. *Physics of the Earth and Planetary Interiors* 2: 19–29.
Hansen BMS (2009) Formation of the terrestrial planets from a narrow annulus. *Astrophysical Journal* 703: 1131–1140.
Hayashi C, Nakazawa K, and Mizuno H (1979) Earth's melting due to the blanketing effect of the primordial dense atmosphere. *Earth and Planetary Science Letters* 43: 22–28.
Hayashi C, Nakazawa K, and Nakagawa Y (1985) In: Black DC and Matthews MS (eds.) *Formation of the Solar System*, pp. 1100–1153. Tucson, AZ: University of Arizona Press.
Herd CDK, Borg LE, Jones JH, and Papike JJ (2002) Oxygen fugacity and geochemical variations in the martian basalts: Implications for martian basalt petrogenesis and the oxidation state of the upper mantle of Mars. *Geochimica et Cosmochimica Acta* 66(11): 2025–2036.
Herzberg C, Feigenson M, Skuba C, and Ohtani E (1988) Majorite fractionation recorded in the geochemistry of peridotites from South Africa. *Nature* 332: 823–826.
Hewins RH and Herzberg C (1996) Nebular turbulence, chondrule formation, and the composition of the Earth. *Earth and Planetary Science Letters* 144: 1–7.
Hofmann AW, Hemond C, Sarbas B, and Jochum KP (2005) Yes, there really is a lead paradox. *EOS, Transactions American Geophysical Union* 86, Fall Meeting Supplement, Abstract V23D-05.
Hofmann AW, Jochum KP, Seufert M, and White WM (1986) Nb and Pb in oceanic basalts: New constraints on mantle evolution. *Earth and Planetary Science Letters* 79: 33–45.
Hofmann AW and White WM (1983) Ba, Rb and Cs in the earth's mantle. *Zeitschrift für Naturforschung* 38: 256–266.
Holland G, Cassidy M, and Ballentine C (2009) Meteorite Kr in Earth's mantle suggests a late accretionary source for the atmosphere. *Science* 326: 1522–1525.
Humayun M and Cassen P (2000) Processes determining the volatile abundances of the meteorites and terrestrial planets. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 3–23. Tucson, AZ: University of Arizona Press.
Humayun M and Clayton RN (1995) Potassium isotope cosmochemistry: Genetic implications of volatile element depletion. *Geochimica et Cosmochimica Acta* 59: 2131–2151.
Hunten DM, Pepin RO, and Walker JCG (1987) Mass fractionation in hydrodynamic escape. *Icarus* 69: 532–549.
Huss GR, Meyer BS, Srinivasan G, Goswami JN, and Sahijpal S (2009) Stellar sources of the short-lived radionuclides in the early solar system. *Geochimica et Cosmochimica Acta* 73: 4922–4945.
Ikoma M and Genda H (2006) Constraints on the mass of a habitable planet with water of nebular origin. *Astrophysical Journal* 648: 696–706.
Jacobsen SB (2005) The Hf–W system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences* 33: 531–570.
Jacobsen SB and Harper CL Jr. (1996) Accretion and early differentiation history of the Earth based on extinct radionuclides. In: Basu A and Hart S (eds.) *Earth Processes: Reading the Isotope Code*, pp. 47–74. Washington, DC: AGU.
Javoy M (1999) Chemical earth models. *Comptes Rendus de l'Académie des Sciences* 329: 537–555, Elseviers SAS.
Jochum KP, Hofmann AW, Ito E, Seufert M, and White WM (1983) K, U and Th in mid-ocean ridge basalt glasses and heat production, K/U and K/Rb in the mantle. *Nature* 306: 431–436.
Jochum KP, Hofmann AW, and Seufert M (1993) Tin in mantle-derived rocks – Constraints on Earth evolution. *Geochimica et Cosmochimica Acta* 57: 3585–3595.
Jones JH and Drake MJ (1986) Geochemical constraints on core formation in the Earth. *Nature* 322: 221–228.
Kamber BS and Collerson KD (1999) Origin of ocean-island basalts: A new model based on lead and helium isotope systematics. *Journal of Geophysical Research* 104: 25479–25491.
Karato S-I and Rama Murthy V (1997) Core formation and chemical equilibrium in the Earth—I. Physical considerations. *Physics of the Earth and Planetary Interiors* 100(1–4): 61–79.
Karner JJ, Sutton SR, Papike JJ, Shearer CK, Jones JH, and Newville M (2006) The application of a new vanadium valence oxybarometer to basaltic glasses from the Earth Moon and Mars. *American Mineralogist* 91(2–3): 270–277.
Kita NT, Huss GR, Tachibana S, Amelin Y, Nyquist LE, and Hutcheon ID (2006) Constraints on the origin of chondrules and CAIs from short-lived and long-lived radionuclides. In: Krot AN, Scott ERD, and Reipurth B (eds.) *Chondrules and the Protoplanetary Disk*, pp. 558–587. San Francisco, CA: Astronomical Society of the Pacific.
Kleine T, Mezger K, Munker C, Palme H, and Bischoff A (2004a) ^{182}Hf–^{182}W isotope systematics of chondrites, eucrites and martian meteorites: Chronology of core formation and early mantle differentiation in Vesta and Mars. *Geochimica et Cosmochimica Acta* 68: 2935–2946.
Kleine T, Mezger K, Palme H, and Münker C (2004b) The W isotope evolution of the bulk silicate Earth: Constraints on the timing and mechanisms of core formation and accretion. *Earth and Planetary Science Letters* 228: 109.
Kleine T, Mezger K, Palme H, Scherer E, and Münker C (2005a) Early core formation in asteroids and late accretion of chondrite parent bodies: Evidence from ^{182}Hf–^{182}W in CAIs, metal-rich chondrites and iron meteorites. *Geochimica et Cosmochimica Acta* 69: 5805–5818.
Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* 418: 952–955.
Kleine T, Palme H, Mezger K, and Halliday AN (2005b) Hf–W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* 310: 1671–1674.
Kleine T, Touboul M, Bourdon B, et al. (2009) Hf–W chronometry and the accretion and early evolution of asteroids and terrestrial planets. *Geochimica et Cosmochimica Acta* 73: 5150–5178.
König S, Münker C, Hohl S, et al. (2011) The Earth's tungsten budget during mantle melting and crust formation. *Geochimica et Cosmochimica Acta* 75: 2119–2136.

Kortenkamp SJ, Kokubo E, and Weidenschilling SJ (2000) Formation of planetary embryos. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 85–100. Tucson, AZ: The University of Arizona Press.

Kortenkamp SJ and Wetherill GW (2000) Terrestrial planet and asteroid formation in the presence of giant planets. *Icarus* 143: 60–73.

Kramers JD and Tolstikhin IN (1997) Two terrestrial lead isotope paradoxes, forward transport modeling, core formation and the history of the continental crust. *Chemical Geology* 139: 75–110.

Kruijer TS, Sprung P, Kleine T, Leya I, Burkhardt C, and Wieler R (2012) Hf–W chronometry of core formation in planetesimals inferred from weakly irradiated iron meteorites. *Geochimica et Cosmochimica Acta* 99: 287–304.

Kwon S-T, Tilton GR, and Grünenfelder MH (1989) In: Bell K (ed.) *Carbonatites: Genesis and Evolution*, pp. 360–387. London: Unwin-Hyman.

Lee D-C and Halliday AN (1996) Hf–W isotopic evidence for rapid accretion and differentiation in the early Solar System. *Science* 274: 1876–1879.

Lee D-C and Halliday AN (1997) Core formation on Mars and differentiated asteroids. *Nature* 388: 854–857.

Lee D-C and Halliday AN (2000a) Hf–W isotopic systematics of ordinary chondrites and the initial $^{182}Hf/^{180}Hf$ of the Solar System. *Chemical Geology* 169: 35–43.

Lee D-C and Halliday AN (2000b) Accretion of primitive planetesimals: Hf–W isotopic evidence from enstatite chondrites. *Science* 288: 1629–1631.

Lee D-C, Halliday AN, Davies GR, Essene EJ, Fitton JG, and Temdjim R (1996) Melt enrichment of shallow depleted mantle: A detailed petrological, trace element and isotopic study of mantle derived xenoliths and megacrysts from the Cameroon line. *Journal of Petrology* 37: 415–441.

Lee D-C, Halliday AN, Leya I, Wieler R, and Wiechert U (2002) Cosmogenic tungsten and the origin and earliest differentiation of the Moon. *Earth and Planetary Science Letters* 198: 267–274.

Lee D-C, Halliday AN, Snyder GA, and Taylor LA (1997) Age and origin of the Moon. *Science* 278: 1098–1103.

Lee T, Shu FH, Shang H, Glassgold AE, and Rehm KE (1998) Protostellar cosmic rays and extinct radioactivities in meteorites. *Astrophysical Journal* 506: 898–912.

Leya I, Halliday AN, and Wieler R (2003) The predictable collateral consequences of nucleosynthesis by spallation reactions in the early Solar System. *Astrophysical Journal* 594: 605–616.

Leya I, Schönbächler M, Wiechert U, Krähenbühl U, and Halliday AN (2008) Titanium isotopes and the radial heterogeneity of the solar system. *Earth and Planetary Science Letters* 266: 233–244.

Leya I, Wieler R, and Halliday AN (2000) Cosmic-ray production of tungsten isotopes in lunar samples and meteorites and its implications for Hf–W cosmochemistry. *Earth and Planetary Science Letters* 175: 1–12.

Li J and Agee CB (1996) Geochemistry of mantle–core differentiation at high pressure. *Nature* 381: 686–689.

Li J and Agee CB (2001) The effect of pressure, temperature, oxygen fugacity and composition on partitioning of nickel and cobalt between liquid Fe–Ni–S alloy and liquid silicate: Implications for the Earth's core formation. *Geochimica et Cosmochimica Acta* 65: 1821–1832.

Liebske C, Corgne A, Frost DJ, Rubie DC, and Wood BJ (2005) Compositional effects on element partitioning between Mg–silicate perovskite and silicate melts. *Contributions to Mineralogy and Petrology* 149: 113–128.

Liew TC, Milisenda CC, and Hofmann AW (1991) Isotopic contrasts, chronology of element transfers and high-grade metamorphism: The Sri Lanka Highland granulites, and the Lewisian (Scotland) and Nuk (S.W. Greenland) gneisses. *Geologische Rundschau* 80: 279–288.

Lissauer JJ (1987) Time-scales for planetary accretion and the structure of the protoplanetary disk. *Icarus* 69: 249–265.

Lissauer JJ (1993) Planet formation. *Annual Review of Astronomy and Astrophysics* 31: 129–174.

Lodders K (2003) Solar system abundances and condensation temperatures of the elements. *The Astrophysical Journal* 591: 1220–1247.

Lugmair GW and Galer SJG (1992) Age and isotopic relationships between the angrites Lewis Cliff 86010 and Angra dos Reis. *Geochimica et Cosmochimica Acta* 56: 1673–1694.

Lugmair GW and Shukolyukov A (1998) Early solar system timescales according to ^{53}Mn–^{53}Cr systematics. *Geochimica et Cosmochimica Acta* 62: 2863–2886.

Magna T, Wiechert U, and Halliday AN (2006) New constraints on the lithium isotope compositions of the Moon and terrestrial planets. *Earth and Planetary Science Letters* 243: 336–353.

Markowski A, Quitté G, Halliday AN, and Kleine T (2006) Tungsten isotopic compositions of iron meteorites: Chronological constraints vs. cosmogenic effects. *Earth and Planetary Science Letters* 242: 1–15.

Marty B (2012) The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth and Planetary Science Letters* 313–314: 56–66.

Marty B, Hashizume K, Chaussidon M, and Wieler R (2003) Nitrogen isotopes on the Moon: Archives of the solar and planetary contributions to the inner solar system. *Space Science Reviews* 106: 175–196.

Marty B and Marti K (2002) Signatures of early differentiation of Mars. *Earth and Planetary Science Letters* 196: 251–263.

McCammon C (1997) Perovskite as a possible sink for ferric iron in the lower mantle. *Nature* 387: 694–696.

McCaughrean MJ and O'Dell CR (1996) Direct imaging of circumstellar disks in the Orion nebula. *Astronomical Journal* 111: 1977–1986.

McDonough WF (2003) Compositional model for the Earth's core. In: Carlson RW (ed.). *The Mantle and Core*, vol. 2, pp. 547–568. Oxford: Elsevier-Pergamon.

McDonough WF and Sun S-S (1995) The composition of the Earth. *Chemical Geology* 120: 223–253.

Merrill PW (1952) Technetium in the Stars. *Science* 115: 484.

Meyer BS and Clayton DD (2000) Short-lived radioactivities and the birth of the Sun. *Space Science Reviews* 92: 133–152.

Meyer BS and Zinner E (2006) Nucleosynthesis. In: Lauretta D, Leshin L, and MacSween H (eds.) *Meteorites and the Early Solar System II*, pp. 69–108. Tucson, AZ: University of Arizona Press.

Mittlefehldt DW, McCoy TJ, Goodrich CA, and Kracher A (1998) Non-chondritic meteorites from asteroidal bodies. In: Papike J (ed.) *Planetary Materials*, *Reviews in Mineralogy*, vol. 36, Ch. 4.

Morbidelli A, Chambers J, Lunine JI, et al. (2000) Source regions and time-scales for the delivery of water to the earth. *Meteoritics and Planetary Science* 35: 1309–1320.

Mostefaoui S, Lugmair GW, Hoppe P, and El Goresy A (2004) Evidence for live Fe-60 in meteorites. *New Astronomy Reviews* 48: 155–159.

Murphy DT, Kamber BS, and Collerson KD (2003) A refined solution to the first terrestrial Pb-isotope paradox. *Journal of Petrology* 44: 39–53.

Murthy VR (1991) Early differentiation of the Earth and the problem of mantle siderophile elements: A new approach. *Science* 253: 303–306.

Newsom HE (1990) Accretion and core formation in the Earth: Evidence from siderophile elements. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 273–288. Oxford: Oxford University Press.

Newsom HE (1995) Composition of the solar system, planets, meteorites, and major terrestrial reservoirs. In: *Global Earth Physics: A Handbook of Physical Constants*. Washington, DC: American Geophysical Union, AGU Reference Shelf, No. 1.

Newsom HE, Sims KWW, Noll PD Jr., Jaeger WL, Maehr SA, and Bessera TB (1996) The depletion of W in the bulk silicate Earth. *Geochimica et Cosmochimica Acta* 60: 1155–1169.

Newsom HE, White WM, Jochum KP, and Hofmann AW (1986) Siderophile and chalcophile element abundances in oceanic basalts, Pb isotope evolution and growth of the earth's core. *Earth and Planetary Science Letters* 80: 299–313.

Nielsen S, Prytulak J, Wood BJ, and Halliday AN (2014) Vanadium isotopic difference between the silicate Earth and meteorites. *Earth and Planetary Science Letters* 389: 167–175.

Nielsen SG, Rehkämper M, and Halliday AN (2006) Large thallium isotopic variations in iron meteorites and evidence for lead-205 in the early Solar System. *Geochimica et Cosmochimica Acta* 70: 2643–2657.

O'Brien DP, Morbidelli A, and Levison HF (2006) Terrestrial planet formation with strong dynamical friction. *Icarus* 184: 39–58.

Ohtani E, Kato T, and Sawamoto H (1986) Melting of a model chondritic mantle to 20 GPa. *Nature* 322: 352–353.

O'Neill HSC (1991a) The origin of the Moon and the early history of the Earth – A chemical model. 1. The Moon. *Geochimica et Cosmochimica Acta* 55(4): 1135–1157.

O'Neill HSC (1991b) The origin of the Moon and the early history of the Earth – A chemical model. 2. The Earth. *Geochimica et Cosmochimica Acta* 55(4): 1159–1172.

O'Neill HSC and Palme H (1998) Composition of the silicate Earth: Implications for accretion and core formation. In: Jackson I (ed.) *The Earth's Mantle – Composition, Structure and Evolution*, pp. 3–127. Cambridge: Cambridge University Press.

O'Neill HSC and Palme H (2008) Collisional erosion and the non-chondritic composition of the terrestrial planets. *Philosophical Transactions of the Royal Society of London, Series A* 366: 4205–4238.

Ozima M and Podosek FA (1999) Formation age of Earth from $^{129}I/^{127}I$ and $^{244}Pu/^{238}U$ systematics and the missing Xe. *Journal of Geophysical Research* 104(B11): 25493–25499.

Pahlevan K and Stevenson DJ (2007) Equilibration in the aftermath of the lunar-forming giant impact. *Earth and Planetary Science Letters* 262: 438–449.

Pahlevan K, Stevenson DJ, and Eiler JM (2011) Chemical fractionation in the silicate vapor atmosphere of the Earth. *Earth and Planetary Science Letters* 301: 433–443.

Palme H (2000) Are there chemical gradients in the inner Solar System? *Space Science Reviews* 92: 237–262.
Patterson CC (1956) Age of meteorites and the Earth. *Geochimica et Cosmochimica Acta* 10: 230–237.
Pepin RO (2000) On the isotopic composition of primordial xenon in terrestrial planet atmospheres. *Space Science Reviews* 92: 371–395.
Pepin RO and Porcelli D (2006) Xenon isotope systematics, giant impacts, and mantle degassing on the early Earth. *Earth and Planetary Science Letters* 250: 470–485.
Pogge von Strandmann PAE, Elliott T, Marschall HR, et al. (2011) Variations of Li and Mg isotope ratios in bulk chondrites and mantle xenoliths. *Geochimica et Cosmochimica Acta* 75: 5247–5268.
Poitrasson F, Halliday AN, Lee D-C, Levasseur S, and Teutsch N (2004) Iron isotope differences between Earth, Moon, Mars and Vesta as possible records of contrasted accretion mechanisms. *Earth and Planetary Science Letters* 223: 253–266.
Porcelli D, Cassen P, and Woolum D (2001) Deep Earth rare gases: Initial inventories, capture from the solar nebula and losses during Moon formation. *Earth and Planetary Science Letters* 193: 237–251.
Porcelli D and Halliday AN (2001) The possibility of the core as a source of mantle helium. *Earth and Planetary Science Letters* 192: 45–56.
Porcelli D and Pepin RO (2000) Rare gas constraints on early earth history. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 435–458. Tucson, AZ: University of Arizona Press.
Pritchard MS and Stevenson DJ (2000) Thermal constraints on the Origin of the Moon. In: Canup R and Righter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tucson, AZ: University of Arizona Press.
Prytulak J, Nielsen S, Ionov DA, et al. (2013) The stable vanadium isotope composition of the mantle and mafic lavas. *Earth and Planetary Science Letters* 365: 177–189.
Pujol M, Marty B, and Burgess R (2011) Chondritic-like xenon trapped in Archean ancient atmosphere. *Earth and Planetary Science Letters* 308: 298–306.
Qin L, Dauphas N, Wadhwa M, Masarik J, and Janney PE (2008) Rapid accretion and differentiation of iron meteorite parent bodies inferred from ^{182}Hf–^{182}W chronometry and thermal modeling. *Earth and Planetary Science Letters* 273: 94–104.
Righter K and Drake MJ (1997) Metal–silicate equilibrium in a homogeneously accreting earth: New results for Re. *Earth and Planetary Science Letters* 146: 541–553.
Righter K and Drake MJ (1999) Effect of water on metal–silicate partitioning of siderophile elements: A high pressure and temperature terrestrial magma ocean and core formation. *Earth and Planetary Science Letters* 171: 383–399.
Righter K, Drake MJ, and Yaxley G (1997) Prediction of siderophile element metal/silicate partition coefficients to 20 GPa and 2800 °C: The effects of pressure, temperature, oxygen fugacity, and silicate and metallic melt compositions. *Physics of the Earth and Planetary Interiors* 100: 115–142.
Ringwood AE (1979) *Origin of the Earth and Moon*. New York: Springer, 295 pp.
Ringwood AE (1989a) Flaws in the giant impact hypothesis of lunar origin. *Earth and Planetary Science Letters* 95: 208–214.
Ringwood AE (1989b) Significance of the terrestrial Mg/Si ratio. *Earth and Planetary Science Letters* 95: 1–7.
Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.
Russell SS, Srinivasan G, Huss GR, Wasserburg GJ, and MacPherson GJ (1996) Evidence for widespread ^{26}Al in the solar nebula and constraints for nebula time scales. *Science* 273: 757–762.
Safronov VS (1954) On the growth of planets in the protoplanetary cloud. *Astronomicheskii Zhurnal* 31: 499–510.
Sasaki S and Nakazawa K (1986) Metal–silicate fractionation in the growing Earth: Energy source for the terrestrial magma ocean. *Journal of Geophysical Research* 91: B9231–B9238.
Savage P, Georg RB, Armytage R, Williams HM, and Halliday AN (2010) Silicon isotope homogeneity in the mantle. *Earth and Planetary Science Letters* 295: 139–146.
Scherstén A, Elliott T, Hawkesworth CJ, Russell S, and Masarik J (2006) Hf–W evidence for rapid differentiation of iron meteorite parent bodies. *Earth and Planetary Science Letters* 241: 530–542.
Schoenberg R and von Blanckenburg F (2006) *Earth and Planetary Science Letters* 252: 342–359.
Schönberg R, Kamber BS, Collerson KD, and Eugster O (2002) New W isotope evidence for rapid terrestrial accretion and very early core formation. *Geochimica et Cosmochimica Acta* 66: 3151–3160.
Seitz H-M, Brey GP, Weyer S, et al. (2006) Lithium isotope compositions of Martian and lunar reservoirs. *Earth and Planetary Science Letters* 245: 6–18.
Shu FH, Shang H, Glassgold AE, and Lee T (1997) X-rays and fluctuating x-winds from protostars. *Science* 277: 1475–1479.
Shukolyukov A and Lugmair GW (1993) Live iron-60 in the early Solar System. *Science* 259: 1138–1142.
Shukolyukov A and Lugmair GW (2004) Manganese–chromium isotope systematics of enstatite meteorites. *Geochimica et Cosmochimica Acta* 68: 2875–2888.
Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–360. Tucson, AZ: University of Arizona Press.
Solomon SC (1979) Formation, history and energetics of cores in the terrestrial planets. *Earth and Planetary Science Letters* 19: 168–182.
Stevenson DJ (1981) Models of the Earth's core. *Science* 214: 611–619.
Stixrude L and Kirko B (2005) Structure and freezing of $MgSiO_3$ liquid in Earth's lower mantle. *Science* 310: 297–299.
Stone JM and Norman ML (1992) The 3-dimensional interaction of a supernova remnant with an interstellar cloud. *Astrophysical Journal* 390: L17.
Suzuki A and Ohtani E (2003) Density of peridotite melts at high pressure. *Physics and Chemistry of Minerals* 30: 449–456.
Takafuji N, Hirose K, Mitome M, and Bando Y (2005) Solubilities of O and Si in liquid iron in equilibrium with (Mg, Fe)SiO_3 perovskite and the light elements in the core. *Geophysical Research Letters* 32: L06313. http://dx.doi.org/10.1029/2005GL022773.
Taylor SR and Norman MD (1990) Accretion of differentiated planetesimals to the Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 29–43. Oxford: Oxford University Press.
Tolstikhin I and Hofmann AW (2005) Early crust on top of the Earth's core. *Physics of the Earth and Planetary Interiors* 148: 109–130.
Touboul M, Kleine T, Bourdon B, Palme H, and Wieler R (2007) Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* 450: 1206–1209.
Truran JW and Cameron AGW (1978) ^{26}Al production in explosive carbon burning. *Astrophysical Journal* 219: 230.
Vanhala HAT and Cameron AGW (1998) Numerical simulations of triggered star formation. 1. Collapse of dense molecular cloud cores. *Astrophysical Journal* 508: 291–307.
Wade J and Wood BJ (2005) Core formation and the oxidation state of the Earth. *Earth and Planetary Science Letters* 236: 78–95.
Walker JCG (1986) Impact erosion of planetary atmospheres. *Icarus* 68: 87–98.
Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475: 206–209.
Walter MJ, Nakamura E, Tronnes RG, and Frost DJ (2004) Experimental constraints on crystallization differentiation in a deep magma ocean. *Geochimica et Cosmochimica Acta* 68: 4267–4284.
Wänke H (1981) Constitution of the terrestrial planets. *Philosophical Transactions of the Royal Society of London* A303: 287–302.
Wänke H and Dreibus G (1986) Geochemical evidence for formation of the Moon by impact induced fission of the proto-Earth. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 649–672. Houston, TX: Lunar Planetary Institute.
Ward WR (2000) On planetesimal formation: The role of collective behavior. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 65–289. Tucson, TZ: University of Arizona Press.
Wasserburg GJ, Busso M, and Gallino R (1996) Abundances of actinides and short-lived nonactinides in the interstellar medium: Diverse supernova sources for the r-processes. *Astrophysical Journal* 466: L109–L113.
Wasserburg GJ, Busso M, Gallino R, and Raiteri CM (1994) Asymptotic giant branch stars as a source of short-lived radioactive nuclei in the solar nebula. *Astrophysical Journal* 424: 412–428.
Wasserburg GJ, Gallino R, and Busso M (1998) A test of the supernova trigger hypothesis with ^{60}Fe and ^{26}Al. *Astrophysical Journal* 500: L189–L193.
Wasserburg GJ, MacDonald F, Hoyle F, and Fowler WA (1964) Relative contributions of uranium, thorium, and potassium to heat production in the Earth. *Science* 143: 465–467.
Wasserburg GJ, Papanastassiou DA, Tera F, and Huneke JC (1977) Outline of a lunar chronology. *Philosophical Transactions of the Royal Society of London* A285: 7–22.
Wasson JT (1985) *Meteorites, Their Record of Early Solar-System History*. New York: W.H. Freeman, 267 pp.
Weidenschilling SJ (2006) Models of particle layers in the midplane of the solar nebula. *Icarus* 181: 572–586.
Wetherill GW (1980) Formation of the terrestrial planets. *Annual Review of Astronomy and Astrophysics* 18: 77–113.
Wetherill GW (1986) Accumulation of the terrestrial planets and implications concerning lunar origin. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 519–550. Houston, TX: Lunar Planetary Institute.

Wetherill GW (1994) Provenance of the terrestrial planets. *Geochimica et Cosmochimica Acta* 58: 4513–4520.

Weyer S, Anbar AD, Brey GP, Münker C, Mezger K, and Woodland AB (2005) Iron isotope fractionation during planetary differentiation. *Earth and Planetary Science Letters* 240: 251–264.

Wiechert U and Halliday AN (2007) Non-chondritic magnesium and the origins of the inner terrestrial planets. *Earth and Planetary Science Letters* 256: 360–371.

Wiechert U, Halliday AN, Lee D-C, Snyder GA, Taylor LA, and Rumble DA (2001) Oxygen isotopes and the Moon-forming giant impact. *Science* 294: 345–348.

Wiechert U, Halliday AN, Palme H, and Rumble D (2004) Oxygen isotopes and the differentiation of planetary embryos. *Earth and Planetary Science Letters* 221: 373–382.

Wilde SA, Valley JW, Peck WH, and Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409: 175–178.

Williams HM, Markowski A, Quitté G, Halliday AN, Teutsch N, and Levasseur S (2006) Iron isotope fractionation in iron meteorites: New insights into metal–sulfide segregation and the sulphur contents of planetary cores. *Earth and Planetary Science Letters* 250: 486–500.

Williams HM, McCammon CA, Peslier AH, et al. (2004) Iron isotope fractionation and oxygen fugacity in the mantle. *Science* 304: 1656–1659.

Williams HM, Wood BJ, Wade J, Frost DJ, and Tuff J (2012) *Earth and Planetary Science Letters* 321–322: 54–63.

Wood BJ (2000) Phase transformations and partitioning relations in peridotite under lower mantle conditions. *Earth and Planetary Science Letters* 174(3–4): 341–354.

Wood BJ, Bryndzia LT, and Johnson KE (1990) Mantle oxidation state and its relationship to tectonic environment and fluid speciation. *Science* 248: 337–345.

Wood BJ and Halliday AN (2005) Cooling of the Earth and core formation after the giant impact. *Nature* 437: 1345–1348.

Wood BJ and Halliday AN (2010) The lead isotopic age of the Earth can be explained by core formation alone. *Nature* 465: 767–770.

Wood BJ and Rubie DC (1996) The effect of alumina on phase transformations at the 660-kilometer discontinuity from Fe–Mg partitioning experiments. *Science* 273(5281): 1522–1524.

Yi W, Halliday AN, Alt JC, et al. (2000) Cadmium, indium, tin, tellurium and sulfur in oceanic basalts: Implications for chalcophile element fractionation in the earth. *Journal of Geophysical Research* 105: 18927–18948.

Yi W, Halliday AN, Lee D-C, and Christensen JN (1995) Indium and tin in basalts, sulfides and the mantle. *Geochimica et Cosmochimica Acta* 59: 5081–5090.

Yin QZ, Jacobsen SB, Yamashita K, Blicher-Toft J, Télouk P, and Albarède F (2002) A short time scale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* 418: 949–952.

Yoshino T, Walter MJ, and Katsura T (2003) Core formation in planetesimals triggered by permeable flow. *Nature* 422: 154–157.

Zahnle K, Arndt N, Cockell C, et al. (2006) Emergence of a habitable planet. *Space Science Reviews*.

Zanda B, Hewins RH, Bourot-Denise M, Bland PA, and Albarède F (2006) Formation of solar nebula reservoirs by mixing chondritic components. *Earth and Planetary Science Letters* 248: 650–660.

Zartman RE and Haines SM (1988) The plumbotectonic model for Pb isotopic systematics among major terrestrial reservoirs – A case for bi-directional transport. *Geochimica et Cosmochimica Acta* 52: 1327–1339.

Zhang J, Dauphas N, Davis AM, Leya I, and Fedkin A (2012) The proto-Earth as a significant source of lunar material. *Nature Geoscience* 5: 251–255.

9.03 Formation of the Earth's Core

DC Rubie, Universität Bayreuth, Bayreuth, Germany
F Nimmo, University of California, Santa Cruz, CA, USA
HJ Melosh, Purdue University, West Lafayette, IN, USA

© 2015 Elsevier B.V. All rights reserved.

9.03.1	**Core Formation in the Earth and Terrestrial Planets**	43
9.03.1.1	Introduction and Present State of Cores in Solar System Bodies	43
9.03.1.2	The Relevance of Iron Meteorites	44
9.03.1.3	History of Early Ideas on Core Formation	45
9.03.2	**Physics of Core Formation**	46
9.03.2.1	Accretion	46
9.03.2.2	Thermal Evolution	47
9.03.2.2.1	Decay of radioactive nuclides	48
9.03.2.2.2	Heating due to the energy of impacts	49
9.03.2.2.3	Heating through the reduction of gravitational potential energy	50
9.03.2.3	Differentiation Mechanisms	51
9.03.2.3.1	Percolation	51
9.03.2.3.2	Metal-silicate separation in a magma ocean	54
9.03.2.3.3	Diapirs and diking	56
9.03.2.3.4	Summary and implications for chemical equilibration	57
9.03.3	**Observational and Experimental Constraints**	58
9.03.3.1	Core Formation Timescales	58
9.03.3.2	Constraints from Siderophile Element Geochemistry	59
9.03.3.2.1	Introduction to siderophile element geochemistry	59
9.03.3.2.2	Core formation/accretion models	61
9.03.3.2.3	Metal–silicate fractionation models	68
9.03.3.2.4	Concluding remarks	70
9.03.3.3	Light Elements in the Core	70
9.03.4	**Summary**	72
Acknowledgments		74
References		74

9.03.1 Core Formation in the Earth and Terrestrial Planets

9.03.1.1 Introduction and Present State of Cores in Solar System Bodies

The Earth's metallic core, comprising 32% of the total mass of the planet, lies beneath a silicate mantle, with the core–mantle boundary (CMB) located at a depth of 2891 km. The differentiation of the Earth into a metallic core and silicate mantle occurred during the accretion of the planet and represents the most important differentiation event in its history. Other terrestrial planets (e.g., Mercury, Venus, and Mars) and small asteroid bodies also underwent such differentiation events during the early history of the solar system and are thus important for providing additional information that helps in understanding differentiation of the Earth.

'Core formation' implies a single event, whereas in reality, the core of the Earth most likely formed through a long series of events, over an extended time period. In this chapter, we consider the period up to which the core reached roughly 99% of its present-day mass; its later evolution is considered in the chapter by Nimmo (Chapter 9.08). Other relevant chapters in this volume are concerned with the early Earth's composition by Halliday and Wood (Chapter 9.02) and the terrestrial magma ocean by Solomatov (Chapter 9.04).

Here, we first provide a general outline of the physical processes that are likely to be involved in core formation, including a discussion of the various uncertainties that arise. The second part of this chapter is focused on observations and experimental data that place constraints on the processes that operated and the state of the core during its formation.

The ultimate reason that the Earth and planets have cores is a matter of elementary physics: Because metallic iron and its alloys are denser than silicates or ices, the most stable arrangement of a rotating, self-gravitating mass of material is an oblate spheroid with the dense iron at the center. Although the physical imperative driving core formation is simple, its realization is complicated by the usual factors of contingency and history. How the cores of the terrestrial planets came to their present configuration depends on what materials were available to make them, at what time and in what condition this material was added, and then how this mass of material evolved with time.

The origin and abundance of the elements of our solar system is now understood as the consequence of (mainly) stellar nucleosynthesis. Nuclear burning in stars created the

elements and established their abundances in our solar system. Iron, as the end product of nuclear burning, owes its large abundance to its maximal nuclear stability. However, this chapter cannot go that far back. Fortunately, there are many good reviews of current ideas about the origin of the elements in our solar system (e.g., Busso et al., 1999; Turan, 1984). But we cannot entirely ignore this aspect of solar system history, because one of the major developments in the past few years follows from the contingent circumstance that the solar system formed in conjunction with one or more nearby supernovas. These catastrophic events produced a substantial abundance of the short-lived radioactive isotopes ^{60}Fe and ^{26}Al, among many others. The consequences of this accident reverberate through our present solar system in ways that are just now becoming clear.

However, before exploring the antecedents of planetary cores in detail, we first consider what is known. All of the terrestrial planets and satellites are believed to possess a dense, probably metallic, core. This belief is founded mainly on the density and moment of inertia of each body. If the average (uncompressed) density of a planet is much larger than the density of the material at its surface, one must infer that its interior is denser. A homogeneous body has a moment of inertia ratio C/MR^2 equal to 0.400. Substantially smaller values of this ratio indicate a central mass concentration, one that usually implies a change of state or composition in the interior (we here exclude the slight decrease in this ratio due to self-compression).

The size of the Earth's core is known accurately from seismic data. The limited operating lifetime of the Apollo lunar seismic array provided tentative evidence for a lunar core (Weber et al., 2011), and as yet no seismic data exist for any other solar system object. Mercury presently possesses a small magnetic field and Mars and the Moon probably had one in the past, suggesting the presence of metallic, fluid, and electrically conducting cores in these planets. Although Venus and Mars do not have magnetic fields at the present time, spacecraft measurements of their k_2 Love number indicate a large deformation in response to solar tides, implying an at least partially liquid core; observations of lunar nutations likewise suggest a liquid lunar core. Some asteroid parent bodies also apparently developed core dynamos (e.g., Fu et al., 2012; Tarduno et al., 2012). **Table 1** summarizes our present knowledge of cores in the terrestrial planets and other bodies in the solar system.

9.03.1.2 The Relevance of Iron Meteorites

During the two centuries after meteorites were first accepted as samples of other bodies in our solar system, approximately 18 000 iron meteorites have been cataloged that, on the basis of trace element groupings, appear to have originated as the cores of approximately 100 separate bodies. These iron meteorite parent bodies were small: cooling rates estimated from Fe/Ni diffusion profiles across taenite/kamacite crystal contacts suggest parent body diameters ranging from 30 to 100 km (Chabot and Haack, 2006; Wasson, 1985). Until a few years ago, it was believed that the iron meteorite parent bodies differentiated into an iron–nickel core and mantle sometime after most of the other meteorites, principally chondrites, had formed. However, dates obtained using the extinct (9 Myr half-life) ^{182}Hf–^{182}W radioactive system have demonstrated that magmatic iron meteorites actually formed about 3 Myr *before* most chondrites (Scherstén et al., 2006). This is nearly the same age as the heretofore oldest objects in the solar system, the CAIs (calcium–aluminum inclusions) found in carbonaceous chondrites. This observation has completely changed our perception of the events in the early solar system and, in particular, the nature of the planetesimals that accumulated to form the Earth and other terrestrial planets.

Table 1 Planetary and satellite cores

Body	*Mean density ($Mg\ m^{-3}$)*	*Moment of inertia factor, C/MR^2*	*Tidal Love number, k_2*	*Mean planet radius, R_p (km)*	*Core radius (km)*	*Magnetic moment, R_p^3 (T)*	*Core mechanical state*	*Composition*
Mercury	5.427	0.346[a]	?	2440	~1600	4×10^{-7}[b]	Liquid	Fe, Ni?
Venus	5.204	~0.33	~0.25	6051.8	~3200	None at present	Liquid	Fe, Ni?
Earth	5.515	0.3308	0.299	6371	3485	6.1×10^{-5}	Liquid outer, solid inner core	Fe, Ni, FeO/FeS?
Moon	3.344	0.3935	0.026[c]	1737.53	400?	None at present	Liquid?	Fe, Ni?
Mars	3.933	0.366	~0.14	3389.9	~1700	Only local sources	Liquid?	Fe, Ni, FeO/FeS?
Io[d]	3.53	0.378	?	1821	~950	None	Liquid?	Fe, Ni, FeS?
Europa[d]	2.99	0.346	?	1565	200–700	None	Liquid?	Fe, Ni, FeS?
Ganymede[d]	1.94	0.312	?	2631	650–900	7.15×10^{-7}[d]	Liquid?	Fe, Ni, FeS?
Callisto[d]	1.83	0.355?	?	2410	None?	None	–	–

[a] Margot et al. (2012).
[b] Russell and Luhmann (1997).
[c] Goossens et al. (2011).
[d] Schubert et al. (2004).
Unless otherwise noted, data are from Yoder (1995) The '?' indicates that the values are unknown.

Previous to this revision in the age of the iron meteorites, the source of the heat necessary to differentiate the iron meteorite parent bodies was obscure: Suggestions ran the gamut from electromagnetic induction in solar flares to lightning in the solar nebula. Although the short-lived radioactive isotopes ^{26}Al (half-life 0.74 Myr) and ^{60}Fe (1.5 Myr) were known to have been present in the early solar system, it was believed that, by the time that the iron meteorite parent bodies formed, these potential heat sources had burned themselves out. However, with the new ages, a new view emerges.

The effectiveness of the heating caused by the decay of a radioactive isotope of a stable element can be gauged from the following formula that gives the maximum temperature rise ΔT occurring in a completely insulated sample of material during the time subsequent to its isolation:

$$\Delta T = \frac{f C_m E_D}{c_p} \quad [1]$$

where c_p is the heat capacity of the material, C_m is the concentration of the stable element in the material, f is the fraction of radioactive isotope at the beginning of the isolation interval, and E_D is the nuclear decay energy released into heat (we do not count the energy of neutrinos emitted during beta decay). Table 2 lists the values of this temperature rise for undifferentiated material of carbonaceous chondrite composition at the time of CAI and iron meteorite formation and at the time that the bulk of the chondrites formed, 3 million years later.

The principal implication of this finding for core formation in the Earth is that many (if not most) of the planetesimals that accumulated to form the major planets possessed metallic cores at the time of their accretion. As yet, the consequences of accretional impacts among partially molten planetesimals are not well studied. However, it is clear that if the planetesimals contributing to the growth of planetary embryos had already formed iron cores, chemical equilibration between iron and silicates initially occurred in a low-pressure regime. In addition, the iron that was added to the growing embryos might not have been in the form of small, dispersed droplets, as required for efficient equilibration.

Just how the iron core of an impacting planetesimal mixes with the existing Earth's surface is presently somewhat unclear. Even the largest meteorite impacts on the present Earth seldom preserve more than a trace of the projectile. In most cases, the projectile, if it can be identified at all, is only revealed by geochemical or isotopic tracers in the rock melted by the impact. The largest intact remnant of an impactor presently known is a 25 cm diameter fragment of LL6 chondrite discovered in the melt sheet of the ca. 70 km diameter Morokweng crater (Maier et al., 2006). The impactor that created the 170 km diameter Chicxulub crater is known only from a few millimeter-sized fragments recovered from a deep-sea core (Kyte, 1998). The projectiles that created these craters are at the low end of the size spectrum we expect for planetesimals: the diameter of the Morokweng impactor was about 4 km or more, while the Chicxulub impactor was probably about 15 km in diameter. Both objects probably impacted near the average velocity for asteroidal impactors on the Earth, about 17 km s^{-1}. Although this velocity is substantially higher than the encounter velocity of the nearly formed Earth with infalling planetesimals (less than about 10 km s^{-1}, or a factor of more than three times less energy than current asteroidal impacts), it still illustrates the fact that impacts typically disrupt the impactor, and whatever the arrangement of materials might have been in the impacting objects, this arrangement is greatly distorted during the impact event.

If we can extrapolate from these observations to the impacts of much larger objects, one would conclude that it makes little difference whether the impacting planetesimal was differentiated or not – in either case, the result of the impact is a finely dispersed mixture of melted target and projectile material. On the other hand, computer simulations of the much larger Moon-forming impact show that the core of a planet-sized impactor remains mostly together in a discrete mass during the impact and appears to merge almost *en masse* with the Earth's core (Canup, 2004). In these simulations, however, each 'particle' representing the material of the Earth and projectile is about 200 km in diameter, so that it is not possible to resolve details of how the iron mass from the projectile core really interacts with the Earth's mantle and core. It thus seems possible that the cores of large planetesimals might remain intact in homogeneous masses too large for chemical equilibration (i.e., larger than a few centimeters) with their surroundings, at least at the beginning of their descent into the Earth's mantle. This is an area needing further study from the impact perspective. Later in this chapter, we discuss the probable fate of such large masses of iron in their inexorable fall toward the Earth's core.

9.03.1.3 History of Early Ideas on Core Formation

Ideas on how the Earth's core formed have shifted dramatically over the past century. At the beginning of the twentieth century, most geophysicists believed, following Lord Kelvin, that the Earth began in a completely molten state and its subsequent history was one of secular cooling and solidification (Thomson and Tait, 1883, p. 482). As late as 1952, Jeffreys

Table 2 Temperature rise of undifferentiated carbonaceous chondrites due to radioactive decay

Radioisotope	*Half-life (Myr)*	*Fractional abundance of isotope at CAI time*, f	*Nuclear decay energy*, E_D *(J kg^{-1})*	*ΔT at CAI time (K)*	*ΔT 3 Myr later (K)*
^{26}Al	0.74	$(5–7) \times 10^{-5}$	1.16×10^{13}[a]	4170	251
^{60}Fe	1.5	4.4×10^{-6}[b]	4.43×10^{12}	2960	740

[a] Schramm et al. (1970)

[b] Quitté et al. (2005). Note that the fractional ^{60}Fe abundance has recently been reestimated as 1.1×10^{-8} (Tang and Dauphas, 2012), which would make this isotope unimportant as a heat source. Assuming the solar system abundances of Anders and Grevesse (1989) and heat capacity $c_p = 1200$ J kg^{-1}K^{-1}. Abundances of Fe and Al are assumed to be chondritic, at 18.2 and 0.865 wt%, respectively (Lodders and Fegley, 1998).

found core formation totally unproblematic, because dense iron would inevitably sink through the liquid protomantle (Jeffreys, 1952, p. 271). However, about the same time, Urey (1952) was elaborating Chamberlin's (1916) hypothesis that the planets had formed from a swarm of small, cold, mutually gravitating 'planetesimals.' In Urey's analysis, core formation becomes highly problematic. The apparent difficulty posed by Urey's view initiated much of our current thinking about how the Earth's metallic iron core originated.

In his famous book *The Planets*, Urey (1952) presented a model of planet formation that strongly appealed to physicists, although, as we shall soon see, it lacked many aspects of reality. Urey approximated the growing Earth as a spherical, homogeneous, isotropic body of radius r that grew from cold matter in space by the addition of infinitesimally thin shells of thickness dr. In this model, he equated the gravitational energy of infalling matter, $-GM(r)/r$ per unit mass, where G is Newton's gravitational constant and $M(r)$ is the mass of the nascent Earth at radius r, with the heat gained after the matter accreted to the Earth. This energy was apportioned between heating of the added mass, heat conduction into the interior, and thermal radiation to space. Because the shell of added matter is very thin, thermal radiation dominates and the planet accretes at very low temperatures.

Models of this kind led Hanks and Anderson (1969) to discover that, even if the Earth accreted in as little as 1 Myr, it would not reach the melting temperature of rock anywhere in its interior. They showed that radioactive heating would only warm the Earth's interior to the melting point much later, initiating core formation as late as 1.6 Gyr after its formation.

By the 1970s, however, study of lead isotopes in ancient crustal rocks indicated that the core had formed within a few hundred million years of the iron meteorites (Gancarz and Wasserburg, 1977) and that a problem existed with the models for thermal evolution of the Earth. Safronov (1978) and Kaula (1979) independently suggested that Urey's model, in which thin shells of infalling matter radiated most of its energy to space, is too drastic. If the impacting planetesimals were of moderate size, a few kilometers or more in diameter, some substantial, but not easily computed, fraction h of the gravitational energy would be buried in the planet. Although this solution has seemed attractive for many decades, it has one central flaw: because the gravitational energy of the planet is initially rather small, the energy added by each unit of mass, $hGM(r)/r$, increases roughly as the square of the radius of the Earth. The center of the Earth thus starts out cold, although a hotter, molten, outer shell eventually develops around it. We thus end up with an Earth that is thermally inside out: cold in the center and hot on the outside.

This apparent conundrum over the initial thermal structure of the Earth led to a series of clever examinations of the stability of a shell of molten, segregated iron in the hot outer portion of the Earth. Elsasser (1963) suggested that a shell of molten iron would push its way to the Earth's center by diapiric instabilities. Later, Stevenson (1981) showed that this instability is even stronger than Elsasser suspected and that the iron would actually fracture the cold kernel of the Earth on a timescale of hours, supposing that such a global iron layer ever had time to form. Recently, more sophisticated models of this type, investigating, for example, the effects of different rheologies, have been published (Lin et al., 2009, 2011).

In our modern era, in which a much more catastrophic view of the Earth's formation reigns (Wetherill, 1985), the problematic initial thermal profile of the Earth is ameliorated by the ability of gigantic impacts to implant heat deep into a growing planet (Melosh, 1990). Deep, strong heating and core formation can be initiated by impacts themselves, given only that they are large and late enough (Tonks and Melosh, 1992). Magma oceans are now seen as an inevitable consequence of the late accretion of planet-scale protoplanets (Tonks and Melosh, 1993). In this era, the problem is not so much how cores form as how, and under what circumstances, iron and silicate may have equilibrated chemically; and how the current inventories of chemical elements in the crusts and mantles of the Earth and planets were established.

9.03.2 Physics of Core Formation

The Earth is the end product of multiple collisions between smaller protoplanets. This process of accretion results in increased temperatures and, ultimately, melting on a planetary scale. As discussed in the succeeding text, differentiation is unavoidable once melting begins; thus, the accretion process is intimately connected to the manner in which the Earth, and its precursor bodies, underwent differentiation and core formation. In this section, our theoretical understanding of the accretion process and its consequences for core formation are discussed; in **Section 9.03.3**, observational and experimental constraints on these processes are outlined.

Earlier reviews and discussions of the processes enumerated here may be found in Stevenson (1989, 1990), Rubie et al. (2003, 2011), Walter and Tronnes (2004), and Wood et al. (2006). The collection of papers edited by Canup and Righter (2000) is also highly recommended as well as the set of papers in *Philosophical Transactions of the Royal Society A (2008) 366*.

9.03.2.1 Accretion

The basic physics of planetary accretion are now reasonably well understood, although many details remain obscure (see Chambers, 2010, for a useful review). Growth of kilometer-sized objects (planetesimals) from the initial dusty, gaseous nebula must have been a rapid process (occurring within approximately 10^3 years), because otherwise the dust grains would have been lost due to gas drag. At sizes >1 km, mutual gravitational interactions between planetesimals become important. Furthermore, because the largest bodies experience the greatest gravitational focusing, they tend to grow at the expense of smaller surrounding objects. This 'runaway growth' phase, if uninterrupted, can potentially result in the development of tens to hundreds of Mars- to Moon-sized embryos in $\sim 10^5$ years at a distance of around 1 astronomical unit (AU) from the Sun (Wetherill and Stewart, 1993). However, runaway growth slows down as the initial swarm of small bodies becomes exhausted and the velocity dispersion of the remaining larger bodies increases (Kokubo and Ida, 1998). Thus, the development of Moon- to Mars-sized embryos probably took $\sim 10^6$ years at 1 AU (Weidenschilling et al., 1997) and

involved collisions both between comparably sized embryos and between embryos and smaller, leftover planetesimals. Based on astronomical observations of dust disks (Haisch et al., 2001), the dissipation of any remaining nebular gas also takes place after a few million years; the dissipation timescale of gas has implications for the orbital evolution of the bodies (e.g., Kominami et al., 2005), their volatile inventories (e.g., Porcelli et al., 2001), and their surface temperatures (e.g., Abe, 1997) and is currently a critical unknown parameter. Noble gas isotopes, in particular those of xenon, have been used to argue for a primordial, dense, radiatively opaque terrestrial atmosphere (e.g., Halliday, 2003; Porcelli et al., 2001), but this interpretation remains controversial (see **Chapter 9.02**).

A recently recognized complication in the process of terrestrial planet formation is that the orbits of the gas giants may have evolved with time, due to scattering of planetesimals. Reorganization of the architecture of the outer solar system may have profoundly affected the inner solar system. For instance, when Jupiter and Saturn encountered the 2:1 resonance, the inner planets may have experienced a potentially prolonged period of high impact flux (Bottke et al., 2012; Gomes et al., 2005). Rapid inward-then-outward migration of Jupiter (the 'Grand Tack' scenario) may have stunted the growth of Mars (Walsh et al., 2011). Gas giant migration was certainly important in the formation of many exoplanetary systems (Chambers, 2009).

Collisional growth processes lead to a peculiar size–frequency spectrum of the accumulating bodies. At first, the runaway accretional processes produce a spectrum in which the cumulative number of objects (the number of objects equal to or greater than diameter D) is proportional to an inverse power of their diameter, generally of form $N_{\rm cum}(D) \sim D^{-b}$, where b is often approximately 2 (Melosh, 1990). One of the principal characteristics of such a distribution is that, although the smallest bodies overwhelmingly dominate in number, most of the mass and energy reside in the very largest objects. Accretional impacts are thus catastrophic in the sense that objects at the largest end of the size spectrum dominate planetary growth. Later, during oligarchic growth at the planetary embryo scale, the large bodies represent an even larger fraction of the size spectrum and giant impacts, that is, impacts between bodies of comparable size, dominate planetary growth history.

The subsequent growth of Earth-sized bodies from smaller Mars-sized embryos is slow, because the embryos grow only when mutual gravitational perturbations lead to crossing orbits. Numerical simulations show that Earth-sized bodies take 10–100 Myr to develop (e.g., Agnor et al., 1999; Chambers and Wetherill, 1998; Morbidelli et al., 2000; Morishima et al., 2010; O'Brien et al., 2006; Raymond et al., 2007) and do so through a relatively small number of collisions between objects of roughly comparable sizes. A result of great importance is that geochemical observations, notably using the Hf–W isotopic system, have been used to verify the timescales obtained theoretically through computer simulations of accretion processes (see **Section 9.03.3.1**).

It should be noted that an important implicit assumption of most late-stage accretion models is that collisions result in mergers. In fact, this assumption is unlikely to be correct (Agnor and Asphaug, 2004; Asphaug, 2010; Asphaug et al., 2006) and many collisions may involve little net transfer of material, though both transient heating and transfer of angular momentum will occur. In fact, nearly 80% of the mantle of Mercury may have been *lost* by collisional erosion after core formation, thus explaining the huge size (mass fraction) of its metallic core (Benz et al., 2007). Such disruptive collisions may also have influenced the bulk chemistry of the terrestrial planets and could explain why some elements (such as Sm–Nd; Caro et al., 2008) are not present in CI chondritic ratios (O'Neill and Palme, 2008). Incorporating incomplete accretion into N-body models represents a major computational challenge, though some progress has been made (Chambers, 2008; Dwyer et al., 2015; Kokubo and Genda, 2010; Leinhardt and Richardson, 2005). Ultimately, such studies may significantly change our picture of the timescales and physical and chemical consequences of accretion.

Figure 1(a) shows a schematic example (obtained by splicing together two different accretion simulations) of how a roughly Earth-mass ($1M_e$) body might grow. Here, the initial mass distribution consists of 11 lunar-mass embryos ($\approx 0.01M_e$) and 900 smaller ($\approx 0.001M_e$) noninteracting planetesimals centered around 1 AU. The solid line shows the increase in mass, and the crosses show the impactor–target mass ratio γ (both in log units). The early stage of growth is characterized by steady collisions with small planetesimals and occasional collisions with other comparably sized embryos (e.g., at 0.068 and 1.9 Myr). Because the planetesimals do not grow, the impactor–target mass ratio γ of colliding planetesimals declines with time; embryo–embryo collisions show up clearly, having $\gamma \sim 1$. At 2 Myr, the growing object has a mass of $0.2M_e$ and roughly half of this mass has been delivered by large impacts. The late stage of growth consists entirely of large impacts, between embryos of comparable masses ($\gamma \sim 0.5$). This final stage takes place over a more extended timescale – in this case, the last significant collision occurs at 14 Myr, resulting in a final mass of $0.73M_e$.

One of the most important outstanding questions regarding this late-stage accretion is the amount of water that was delivered to the Earth. The presence of large quantities of water in the early mantle would have profound implications for the oxidation state and composition of the core (see Williams and Hemley, 2001); furthermore, a by-product would be a thick steam atmosphere, which would be sufficiently insulating to ensure a magma ocean (Matsui and Abe, 1986). Although the Earth formed inside the 'snow line,' where water ice becomes unstable, its feeding zone probably expanded over time and some of its constituent planetesimals were likely derived from greater heliocentric distances and thus contained more water. Simulations (Morbidelli et al., 2000; O'Brien et al., 2006; Raymond et al., 2007) suggest that a water-rich Earth is quite likely, but the stochastic nature of the outcomes precludes a firm conclusion. Radial mixing of planetesimals is clearly not completely efficient because of the differing oxygen isotope characteristics of Earth and Mars (e.g., Clayton and Mayeda, 1996).

9.03.2.2 Thermal Evolution

As discussed in the succeeding text in **Section 9.03.2.3**, the actual mechanisms of core formation (metal–silicate

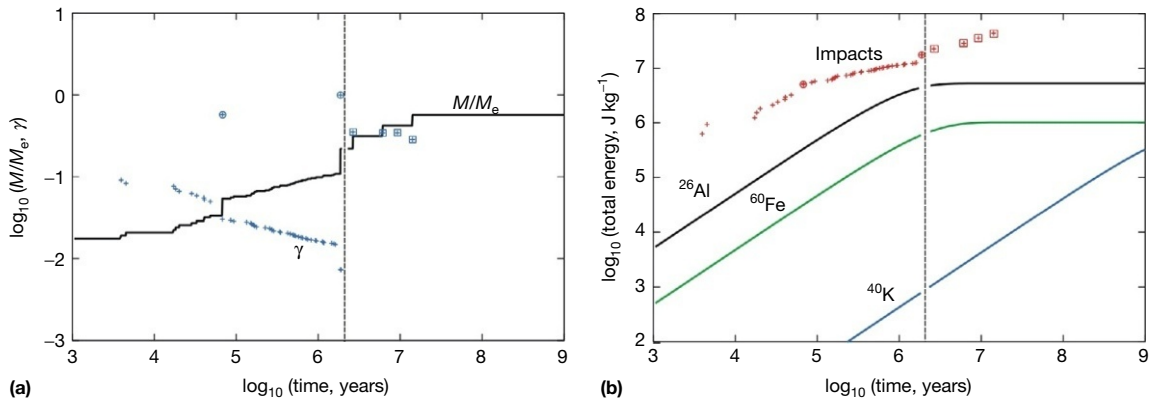


Figure 1 (a) Schematic growth of a proto-Earth, obtained by splicing two accretion simulations together. Early growth is from Agnor (unpublished) where the initial mass distribution consists of 11 embryos ($\approx 0.01 M_e$) and 900 noninteracting planetesimals ($\approx 0.001 M_e$) centered around 1 AU. Late growth is from particle 12 in run 3 of Agnor et al. (1999). The vertical dashed line denotes the splicing time. The solid line shows the mass evolution of the body, and the crosses denote the impactor–target mass ratio γ. Circles denote embryo–embryo collisions; squares denote late-stage giant impacts. The general reduction in γ prior to 2 Myr is a result of the fact that the planetesimals cannot merge with each other, but only with embryos. (b) Corresponding energy production (J kg^{-1}). The cumulative energy due to impacts (crosses) is calculated using eqn [2] for each impact. The solid lines show the cumulative energy associated with the decay of radioactive elements ^{26}Al, ^{60}Fe, and ^{40}K. Half-lives are 0.73 Myr, 1.5 Myr, and 1.25 Gyr, respectively; initial bulk concentrations are 5×10^{-7}, 2×10^{-7}, and 4.5×10^{-7}, respectively (Ghosh and McSween, 1998; Tachibana et al., 2006; Turcotte and Schubert, 2002).

separation) that operate are dependent on the thermal state of a planetary body and at least some degree of partial melting is required. However, in addition to understanding the thermal state of the Earth during core formation, it is also important to understand the thermal histories of small bodies (planetesimals and asteroids) because these determine whether or not the material that accreted to form the Earth had already undergone core–mantle differentiation.

There are three main sources of energy that can produce the melting that is required for core formation. Firstly, the decay of short-lived radioactive nuclides (^{26}Al and ^{60}Fe) is an important source of energy when accretion occurs very soon after the formation of the solar system (**Figure 1(b)**). (These isotopes have half-lives of 0.73×10^6 and 1.5×10^6 years, respectively.) Secondly, the kinetic energy delivered by impacts can be sufficient to generate local or global melting especially during the late stages of Earth accretion (**Figures 1(b)** and **3**). Finally, as discussed in the succeeding text, the process of differentiation itself, by reducing the gravitational potential energy of the body, also releases heat and may lead to runaway differentiation.

9.03.2.2.1 Decay of radioactive nuclides

Thermal models show that the decay of ^{26}Al and ^{60}Fe in a body with a minimum radius of 30 km can result in maximum temperatures that range from below the Fe–FeS eutectic temperature to above silicate melting temperatures, depending on the initial concentrations of these isotopes (**Figure 2**). In contrast, the energy released through collisions between bodies less than a few hundred km radius is insufficient to cause global melting (Keil et al., 1997; see also **Section 9.03.2.2.2**). This means that the melting required for core–mantle differentiation in a small body could only occur at a very early stage during the history of the solar system – for example, within the first 1 Myr (Baker et al., 2005). In support of the thermal models, there is geochemical evidence for large-scale melting

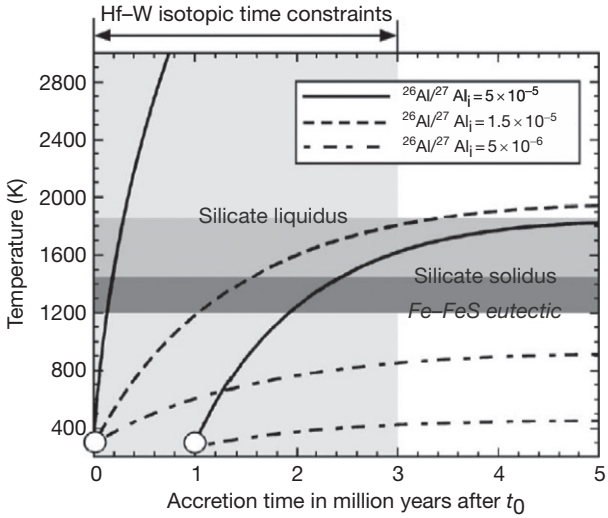

Hf–W isotopic time constraints
26Al/27 Al_i = 5 × 10−5
26Al/27 Al_i = 1.5 × 10−5
26Al/27 Al_i = 5 × 10−6
Silicate liquidus
Silicate solidus
Fe–FeS eutectic
Temperature (K)
Accretion time in million years after t0

Figure 2 Models of the thermal evolution of a small body (>30 km radius) as a function of the initial concentration of ^{26}Al. Temperature is calculated as a function of time after the start of the solar system (t_0). Depending on the concentrations of ^{26}Al and ^{60}Fe and when accretion starts, maximum temperatures range from below the Fe–FeS eutectic to above the silicate solidus (see also Yoshino et al., 2003, **Figure 2**). Reproduced from Walter MJ and Tronnes RG (2004). Early Earth differentiation. *Earth and Planetary Science Letters* 225: 253–269.

and magma ocean formation on at least some small bodies (Greenwood et al., 2005). In addition, the parent body of the HED meteorites (which is likely asteroid 4 Vesta, 530 km in diameter) underwent early core–mantle differentiation. These considerations support the view that planetesimals that accreted to form the Earth were already differentiated (e.g., Taylor and Norman, 1990). However, if small bodies accreted somewhat later, melting and differentiation may not have occurred because the concentrations of ^{26}Al and ^{60}Fe would

have been too low (Figure 2). The density of the solar nebula must have decreased with increasing distance from the Sun (e.g., Figure 1 in Elser et al., 2012). Because the rate of a body's accretion depends on the density of the solar nebula, bodies in the outer regions would have accreted slower and therefore later than those close to the Sun (Grimm and McSween, 1993). Thus, not all planetesimals would have differentiated. Ceres is an example of an asteroid that may not have differentiated to form a metallic core (Thomas et al., 2005). It is therefore currently not clear whether the Earth accreted from only differentiated bodies or a mixture of differentiated and undifferentiated bodies. The answer depends on the extent of the Earth's feeding zone during accretion.

9.03.2.2.2 Heating due to the energy of impacts

Figure 1(a) shows that the bulk of late-stage Earth accretion involves large impacts well-separated in time. The energetic consequences of such impacts have been discussed elsewhere (Benz and Cameron, 1990; Melosh, 1990; Tonks and Melosh, 1993) and strongly suggest that, even in the absence of a thick primordial (insulating) atmosphere, the final stages of Earth's growth must have involved of one or more deep global magma oceans. This conclusion has important implications for the mode of core formation and may be understood using the following simple analysis.

For an impact between a target of mass M and an impactor of mass γM, the mean change in energy per unit mass of the merged object due to kinetic and gravitational potential energy is

$$\Delta E = \frac{1}{1+\gamma}\left[-\frac{3}{5}\left(\frac{4\pi\rho}{3}\right)^{1/3} GM^{2/3}\left(1+\gamma^{5/3}-(1+\gamma)^{5/3}\right)+\frac{1}{2}\gamma V_\infty^2\right] \quad [2]$$

Here, ρ is the mean density of the merged object, G is the universal gravitational constant, V_∞ is the velocity of the impactor at a large distance from the target, and the factor of 3/5 comes from considering the binding energy of the bodies (assumed uniform) prior to and after the collision. Neglecting V_∞ and taking γ to be small, the global average temperature rise associated with one such impact is given by

$$\Delta T \approx 6000K\left(\frac{\gamma}{0.1}\right)\left(\frac{M}{M_e}\right)^{2/3} \quad [3]$$

where we have assumed that $\rho = 5000\ \mathrm{kg\ m^{-3}}$ and a heat capacity of 1 kJ $\mathrm{kg^{-1}\ K^{-1}}$. Note that this temperature change is a globally averaged value; for small impacts in particular, real temperatures will vary significantly with distance from the impact site.

Equation [2] is based on several simplifying assumptions. It assumes that the energy is deposited uniformly, which is unlikely to be correct. More importantly, it assumes that all the kinetic energy is converted into heat and retained. Such an assumption is unlikely to be correct for small impacts, where most of the energy is deposited at shallow depths where it can be radiated back to space (Stevenson, 1989). For larger impacts, however, the energy will be deposited at greater depths, and thus, the only major energy loss mechanism is the ejection of hot material. The amount and temperature of material ejected depend strongly on the geometry of the impact but are in general rather small compared to the target mass (Canup et al., 2001). Since we are primarily concerned with large impacts ($\gamma > 0.1$), the assumption that the majority of the energy is retained as heat energy is a reasonable one. Thus, impactors with a size similar to the one that is believed to have formed the Moon (Cameron, 2000; Canup and Asphaug, 2001) probably resulted in the bulk of the Earth being melted.

Although a Mars-sized ($0.1M_e$) proto-Earth has a smaller mass, it experiences collisions with bodies comparable in size to itself ($\gamma \approx 1$, see Figure 1) In this case, eqn [2] shows that $\Delta T \approx 4500\ K$. Thus, it seems likely that Mars-sized embryos were also molten and thus differentiated. Although there is currently little direct evidence for an ancient Martian magma ocean (see Elkins-Tanton et al., 2003), Debaille et al. (2007), Blichert-Toft et al. (1999), and Borg and Draper (2003) have used Sm–Nd, Lu–Hf, and incompatible element abundances, respectively, to argue for such an ocean. Dauphas and Pourmand (2011) used Hf–W data to argue for a very early formation time for Mars, in which case global melting due to ^{26}Al would have resulted; however, their results are based on assumptions concerning core–mantle reequilibration which may not be correct (see succeeding text).

In considering the thermal effects of impacts, we can consider the cumulative energy delivered in comparison with other sources of energy. Figure 1(b) shows the cumulative impact energy in joule per kilogram. The bulk of the energy is delivered by the few largest impacts, as expected. For comparison, the radioactive heat production due to one long-lived (^{40}K) and two short-lived isotopes (^{26}Al and ^{60}Fe) is shown. Long-lived isotopes have no effect at all on the thermal evolution of the Earth over its first 100 Myr. The total energy associated with ^{26}Al depends very strongly on the accretion time (Figure 2) and in this case is roughly one order of magnitude smaller than that due to the impacts. Figure 1(b) shows that the thermal evolution of the Earth naturally divides into two stages: the early stage (up to ~100 Myr) when heating due to impacts and short-lived isotopes dominate and the later stage when long-lived isotopes and secular cooling are important.

Figure 3 summarizes the expected mean global temperature change due to impacts and short-lived radionuclides as a function of planetary size. The effect of a single impact (solid lines) is calculated using eqn [2] and demonstrates the strong dependence on both the body mass and the impactor–target mass ratio γ. It should be reemphasized that, particularly for small impacts, the energy will not be distributed evenly and that the calculated temperature rise is only a mean global value. The effect of ^{26}Al decay (dashed lines) does not depend on the body mass, but only on the accretion time relative to CAI formation. For small bodies, only radioactive decay contributes substantially to warming; for large bodies, gravitational energy release will tend to dominate. Figure 3 also emphasizes that the population of impactors a growing planet encounters has a very important effect on its ultimate temperature structure. If all the impactors are small ($\gamma \ll 1$), then the temperature changes are quite modest. Conversely, the later stages of planetary accretion ensure that growing bodies will encounter other bodies of similar size (i.e., $\gamma \sim 1$; Section 9.03.2.1), and thus, melting is expected for bodies of roughly Moon size and larger.

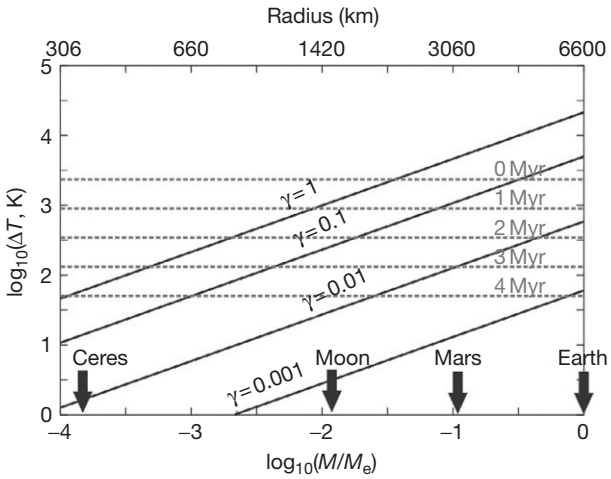


Figure 3 Mean global temperature change ΔT as a function of planetary mass (in units M_e = one Earth mass). Solid lines show gravitational energy due to a single collision where y denotes the impactor–target mass ratio and ΔT is calculated from eqn [2] assuming that $V_\infty = 0$, $C_p = 1000$ J kg^{-1} K^{-1}, and $\rho = 5000$ kg m^{-3}. Dashed lines show temperature change due to ^{26}Al decay as a function of (instantaneous) planet formation time (in Myr) after solar system formation. Total energy release by ^{26}Al is 6.2×10^6 J kg^{-1} (assuming 1 wt% Al and a fractional abundance of ^{26}Al of 5×10^{-5}) and half-life is 0.73 Myr. Planetary melting is expected to be widespread for $\Delta T > 1000$ K.

Although the average temperature jumps in large planets struck by comparably sized objects are impressive, it must be kept in mind that these are averages only. Impacts are not good at homogenizing temperatures. An impact initially deposits a large fraction of its kinetic energy in a region comparable in size to the projectile itself. Shortly after a time measured by D/v_i the projectile diameter D divided by its impact velocity v_i, the shock waves generated by the impact expand away from the impact site, accelerating the target material and depositing heat as they spread out and weaken. The heat is therefore mostly deposited in a region called the 'isobaric core' whose shape is initially approximately that of an immersed sphere tangent to the surface of the target. Outside of this region, the shock pressure, and therefore the amount of heat deposited, falls off rapidly, generally as more than the cube of the distance from the impact (Pierazzo et al., 1997). A velocity field that is established by the spreading shock wave accompanies this heat deposition. The moving material generally opens a crater whose size and form depend on the relative importance of gravity and strength in the crater excavation. In general, about half of the hot, isobaric core is ejected from the impact site, to spread out beyond the crater rim, while the other half remains in the distorted target beneath the crater floor (i.e., for impacts that do not vaporize a large volume of the target, which is the case for all accretional impacts on Earth-sized planets). The major part of the kinetic energy of the impact is thus converted to heat in a roughly hemispheric region centered on the crater. This hot zone extends to a depth of only a few projectile diameters.

For small impacts, generally those of objects less than a kilometer in diameter, the heat is deposited so close to the surface that a major fraction is radiated to space before the next similar-sized impact occurs (Melosh, 1990). Larger impactors deposit more of their energy deeper in the planet, up to the truly gigantic impacts of objects comparable in size to the growing Earth, which may deposit their energy over an entire hemisphere. Nevertheless, detailed computations show that this energy is not homogeneously distributed over the target Earth. Later processes, such as the reimpact of large fractions of the projectile (seen in some of the Moon-forming scenarios: Canup, 2004) or thermal convection of the mantle driven by a suddenly heated core, are needed to homogenize the heat input. Impact heating is thus characterized by large initial temperature variations: The part of the Earth near the impact site becomes intensely hot, while more distant regions remain cold. Averages, therefore, tell only a small part of the overall story: The aftermath of heating by a large impact is characterized by strong temperature gradients, and the evolution of the planet may be dominated by contrasts between very hot and cold, mostly unaffected, portions of the growing Earth.

9.03.2.2.3 Heating through the reduction of gravitational potential energy

Although the energies involved in late-stage impacts imply the formation of at least a regional magma ocean, such an ocean may not reach the CMB, because the solidus temperature of mantle material increases as a function of pressure (see **Figure 9(b)**). The mechanical properties of the magma ocean, which control the rate of iron transport through the ocean, change dramatically when the melt fraction drops below ≈60% (Solomatov, 2000; see also **Chapter 9.04**). The effective base of the magma ocean occurs at this rheological transition. Descending iron droplets will tend to pond at this interface. However, the resulting iron layer is still denser than the underlying (mantle) material and will therefore tend to undergo further transport toward the center of the planet. The transport mechanism might be percolation through a partially molten mantle or the motion of larger iron bodies via brittle fractures (diking) or through a viscously deformable mantle (diapirism). These different mechanisms are discussed in Stevenson (1990) and will be addressed in more detail in the succeeding text (**Section 9.03.2.3**).

The redistribution of mass involved with the descent of iron toward the center of the planet results in a release of gravitational energy. Because the gravitational energy change only depends on the initial and final density distributions, the mechanism by which the iron is transported is of only secondary importance. The extent to which the iron, rather than the surrounding mantle material, is heated depends on the rate of transport and relative viscosities of the iron and silicates (Golabek et al., 2008; Monteux et al., 2009; Samuel et al., 2010), but the total energy released would be the same. The magnitude of the heating can be large and may be calculated as follows.

Consider a uniform, thin layer of iron at the base of the magma ocean, overlying a mantle and core. The top and bottom of the iron layer and the underlying core are at radii R_o, $R_m = (1-\varepsilon) R_o$, and $R_c = \beta R_o$, respectively, where ε is a measure of the thickness of the iron layer ($\varepsilon \ll 1$) and β is a measure of the initial core radius. After the removal of iron to the center, whether by diapirism, diking, or percolation, the core will have grown and the new configuration will have a lower potential

energy. The difference in potential energy may be calculated and used to infer the mean temperature change in the final core, assuming that all the potential energy is converted to core heat (e.g., Solomon, 1979). For the specific case of a constant core density twice that of the mantle density, it may be shown that the mean temperature change of the entire postimpact core is given by

$$\Delta T = \frac{\pi G \rho_c R_o^2}{\beta^2 C_p \left(1 + 3\varepsilon\beta^{-2}\right)} \left[\frac{1}{10}\beta^5 \left(\left(1 + 3\varepsilon\beta^{-3}\right)^{5/3} - 1\right) + \frac{1}{2}\varepsilon - \beta^3\varepsilon\right] \quad [4]$$

Here, C_p is the core-specific heat capacity and ρ_c is the core density, and it is assumed that the core is well mixed (isothermal). As before, the temperature change is a strong function of planetary size, specifically the radial distance to the base of the magma ocean R_o. The temperature change also depends on ε, which controls the mass of iron being delivered to the core, and the initial core radius R_c when $\beta > 0$. The temperature change goes to zero when $\beta = 1$, as expected, while when $\beta = 0$ (i.e., no initial core), the temperature change is essentially independent of the mass of iron delivered. For Earth-sized planets, the addition of iron to the core by individual impacts can lead to core temperature increases of several hundred to a few thousand kelvin. However, this temperature change is generally small compared to that induced by the impact itself, though it may have important effects, for example, regarding the generation of early core dynamos.

For example, consider two cases, appropriate to the Moon-forming impact (Canup and Asphaug, 2001): a $0.9M_e$ planet hit by a $0.1M_e$ impactor and a $0.8M_e$ planet hit by a $0.2M_e$ impactor, all bodies having a core of density 10^4 kg m^{-3} and radius half the body radius. For magma oceans of depths 500 and 1000 km, respectively ($R_o = 5900$ and 5400 km), we obtain $\varepsilon = 0.0054$ and 0.0139 and $\beta = 0.52$ and 0.55. The core adiabat is roughly 1 K km^{-1}, giving temperature increases of 2400 and 1900 K. The further increases from gravitational heating (eqn [4]) are 1250 and 2000 K, respectively. Thus, the postimpact core temperature is likely to have increased by 3500–4000 K from the temperature it attained at the base of the magma ocean. Because the base of the magma ocean is estimated to be in the range 2500–4000 K (**Section 9.03.3.2**), the initial core temperature was probably at least 6000 K, sufficient to cause substantial lower-mantle melting.

This estimate is only approximate, because of the assumptions made (e.g., no transfer of heat to the mantle) and the fact that the Earth probably suffered several comparably sized impacts. However, the result is important because the initial temperature contrast between the core and the lowermost mantle determines the initial CMB heat flux and thus the ability of the core to generate a dynamo (see **Chapter 9.08**).

9.03.2.3 Differentiation Mechanisms

Crucial questions are as follows: at what stage during the accretion history did core–mantle differentiation actually occur and how long did the process take? These questions depend on the thermal history of the accreting body, which, in turn, determines the physical mechanisms by which metal and silicate separate. The physics of differentiation have been reviewed by Stevenson (1990) and Rushmer et al. (2000), and here we provide an updated account. Differentiation occurs because of the large density contrast between silicates and metal, but the rate at which it occurs depends on the length scales and properties of the phases involved. Considering the current size of the Earth, the segregation process involved the transport of metal through the protomantle over length scales of up to almost 3000 km. As reviewed by Stevenson (1990), the process involved the transport of liquid metal through either solid (crystalline) silicates or partially or fully molten silicate (i.e., in a magma ocean). The former mechanism is possible because iron (plus alloying elements) has a lower melting temperature than mantle silicates, so that liquid iron can coexist with solid silicates. In contrast, the separation of solid metal from solid silicate is too sluggish to have been a significant process during core formation (Stevenson, 1990). Identifying whether metal separated from solid or molten silicates during core formation clearly provides information about the early thermal state of the planet. In addition, the separation mechanism may have affected the conditions and extent of chemical equilibration between metal and silicate and therefore affects mantle geochemistry, as discussed further in the succeeding text. Here, we review results pertaining to core formation by (a) grain-scale percolation of liquid metal through crystalline silicates, (b) separation of molten metal from molten silicate in a magma ocean, and (c) descent of large (km-scale) diapirs of molten metal though crystalline silicate.

As discussed in the preceding text, it seems likely that the final stages of Earth's accretion involved large impacts between previously differentiated objects that were at least partly molten. What happens in detail during these impacts is poorly understood. Hydrocode simulations of impacts (Cameron, 2000; Canup and Asphaug, 2001) show that the cores of the target and impactor merge rapidly, within a few free-fall timescales (hours), although a small fraction (typically less than 1%) of core material may be spun out into a disk. Unfortunately, the resolution of these simulations is on the order of 100 km, while the extent to which chemical reequilibration occurs depends on length scales that are probably on the order of centimeters (Stevenson, 1990). Another approach to understanding this important question is presented in the succeeding text in **Section 9.03.2.3.2**.

9.03.2.3.1 Percolation

Liquid metal can percolate through a matrix of polycrystalline silicates by porous flow provided the liquid is interconnected and does not form isolated pockets (Rushmer et al., 2000; Stevenson, 1990). The theory is well developed for the ideal case of a monomineralic system consisting of crystals with isotropic surface energy (i.e., no dependence on crystallographic orientation). Whether the liquid is interconnected depends on the value of the wetting or dihedral angle (θ) between two solid–liquid boundaries that are intersected at a triple junction by a solid–solid boundary (**Figure 4**) (Stevenson, 1990; von Bargen and Waff, 1986). For $\theta < 60°$, the liquid is fully interconnected and can percolate through the solid irrespective of its volume fraction; under such conditions, complete metal–silicate segregation can occur efficiently by porous flow. In the case that $\theta > 60°$, the liquid forms isolated

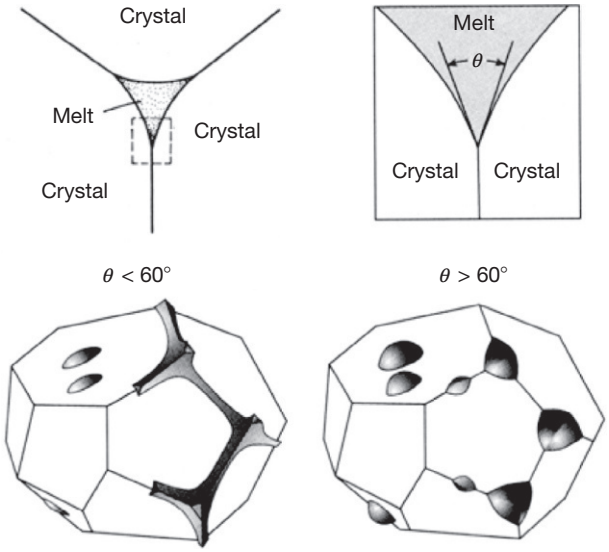


Figure 4 Relation between melt connectivity and dihedral angle in a polycrystalline aggregate containing a small amount of dispersed melt. The dihedral angle is defined in the top diagrams, and the dependence of melt connectivity on the dihedral angle is shown in the lower diagrams. Note that an ideal case is shown here in which the crystals have an isotropic surface energy. Reproduced from Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*. Oxford: Oxford University Press.

pockets when the melt fraction is low and connectivity exists only when the melt fraction exceeds a critical value. This critical melt fraction is known as the connection boundary and ranges from 0% to 6% for dihedral angles in the range 60–85°. If the melt fraction lies above the connection boundary, the melt is interconnected and can percolate. However, as the melt fraction decreases due to percolation, a pinch-off melt fraction is reached below which interconnectivity is broken. At this point, the remaining melt is stranded in the crystalline matrix. The pinch-off melt fraction lies slightly below the connection boundary and ranges from 0% to 4.5% for dihedral angles in the range 60–85°. The dependence of the pinch-off boundary on dihedral angle has been formulated theoretically as 0.009 $(\theta - 60)^{0.5}$ for values of θ in the range 60–100° (von Bargen and Waff, 1986). Yoshino et al. (2003) have shown, through electrical conductivity measurements, that the percolation threshold for Fe–S melt in an olivine matrix is approximately 5 vol.%. This result suggests that core formation could have occurred by percolation under hydrostatic conditions (see succeeding text), but ~5 vol.% metal would have been stranded in the mantle. In fact, 5 vol.% may be a lower bound on the amount of stranded metal because the percolation threshold theory was developed assuming that the liquid is uniformly and finely dispersed throughout the crystalline matrix (von Bargen and Waff, 1986). However, due to the minimization of surface energy, textures might evolve over long time periods (relative to normal experimental timescales of hours to days) such that the liquid becomes concentrated in pools that are widely dispersed and relatively large, in which case the percolation threshold could be much greater than 5 vol.% (Bagdassarov et al., 2009a,b; Stevenson, 1990; Walte et al., 2007).

The effects of crystal anisotropy and crystal faceting are not taken into account by the aforementioned theory (which is based on surface energies being isotropic). It has been argued that the effects of crystal anisotropy and crystal faceting reduce permeability (Faul, 1997; Laporte and Watson, 1995; Yoshino et al., 2006); however, there is no experimental evidence to suggest that such effects are significant for liquid metal–silicate systems relevant to core formation.

The dihedral angle θ depends on the energies of the respective interfaces that intersect at a triple junction that is occupied by a melt pocket (**Figure 4**):

$$\theta = 2\cos^{-1}\left(\frac{\gamma_{ss}}{2\gamma_{sl}}\right) \quad [5]$$

where γ_{sl} is the solid–liquid interfacial energy and γ_{ss} is the solid–solid interfacial energy. Note that, as emphasized previously, this expression is based on the assumption that interfacial energies are independent of crystal orientation and that stress is hydrostatic. When considering metal–silicate systems that are applicable to core formation, the interfacial energies, and therefore the dihedral angle, can be affected by (a) the structure and composition of the crystalline phase, (b) the structure and composition of the liquid metal alloy, and (c) temperature and pressure. Dihedral angles in metal–silicate systems relevant to core formation and the effects of the aforementioned variables have been investigated experimentally (Ballhaus and Ellis, 1996; Gaetani and Grove, 1999; Holzheid et al., 2000a; Minarik et al., 1996; Rose and Brenan, 2001; Shannon and Agee, 1996, 1998; Takafuji et al., 2005; Terasaki et al., 2005, 2007, 2008). These studies, performed on a range of different starting materials at pressure–temperature conditions up to those of the lower mantle, now enable the most important factors that control dihedral angles in metal–silicate systems to be identified.

The effects of pressure, temperature, and the nature of the silicate crystalline phase appear to be relatively unimportant, at least up to ~23 GPa. For example, through experiments on the Homestead meteorite, Shannon and Agee (1996) found that dihedral angles have an average value of 108° and remain essentially constant over the pressure range 2–20 GPa, irrespective of the dominant silicate mineral (e.g., olivine or ringwoodite). However, their subsequent study of the Homestead meteorite under lower-mantle conditions (25 GPa) suggests that dihedral angles decrease to ~71° when silicate perovskite is the dominant silicate phase (Shannon and Agee, 1998). Since this angle is still in excess of 60°, efficient metal segregation from crystalline perovskite under lower-mantle pressures is not possible (see also Terasaki et al., 2007).

The most important parameter controlling dihedral angles is evidently the anion (oxygen and/or sulfur) content of the liquid metal phase. The reason is that dissolved O and S act as 'surface-active' elements in the metallic melt and thus reduce the solid–liquid interfacial energy (e.g., Iida and Guthrie, 1988). Dissolved oxygen, in particular, makes the structure of the metal more compatible with that of the adjacent silicate, thus reducing the interfacial energy and promoting wetting.

Rushmer et al. (2000, their Figure 4) showed that dihedral angles decrease from 100–125° to 50–60° as the anion-to-cation ratio, defined as $(O+S)/(Fe+Ni+Co+Mn+Cr)$, increases from ~0.3 to ~1.2. Based on their data compilation,

only metallic melts with the highest contents of O and S are wetting ($\theta < 60°$). The effects of the anion content of the metallic liquid have been investigated systematically in the stability fields of olivine and ringwoodite by Terasaki et al. (2005) through a study of dihedral angles in the system Fe–S–$(Mg,Fe)_2SiO_4$ in the pressure range 2–20 GPa. By varying the FeO content of the silicate phase (Fe# = FeO/(FeO + MgO) = 0.01–0.44), the oxygen fugacity, which controls the O content of the metal, could be varied over a wide range. They confirmed the importance of the anion content of the liquid phase and showed that the effect of dissolved oxygen is greater than that of dissolved S (**Figure 5**). Because the oxygen content of the metallic liquid decreases with pressure up to 10–15 GPa (Asahara et al., 2007; Frost et al., 2010; Rubie et al., 2004), dihedral angles increase with increasing pressure in this range. Terasaki et al. (2008) showed that for metallic liquids with a high O + S content, dihedral angles lie below 60° at pressures below 2–3 GPa, depending on the FeO content of the silicate phase. Percolation could therefore contribute significantly to core formation in planetesimals during the early stages of heating before the temperatures reach the silicate solidus.

Based on a study of liquid Fe + silicate perovskite at 24–47 GPa and 2400–3000 K, Takafuji et al. (2004) have suggested that dihedral angles decrease to ~51° under deep mantle conditions. This result may be consistent with results that show that the solubility of oxygen in liquid Fe increases strongly with temperature and weakly with pressure above 10–15 GPa (Asahara et al., 2007; Frost et al., 2010). However, the study of Takafuji et al. (2004) was performed using a laser-heated diamond anvil cell (LH-DAC), in which samples are exceedingly small and temperature gradients are very high. Furthermore, dihedral angles had to be measured using transmission electron microscopy – a technique that is far from ideal for obtaining good statistics. This preliminary result therefore awaits confirmation from further studies. In addition, in a study of wetting in a similar system at the lower pressures of 25 GPa, Terasaki et al. (2007) found that the dihedral angle increases with the $FeSiO_3$ component of silicate perovskite but failed to find any obvious correlation with the oxygen content of the metal liquid.

In summary, experimental results for systems under hydrostatic stress show that dihedral angles significantly exceed the critical angle of 60° at pressures of 3–25 GPa in chemical systems that are relevant for core formation in terrestrial planets. This means that, for percolation under static conditions, at least several vol% metal would have been stranded in the mantle – which is inconsistent with the current concentration of siderophile elements in the mantle (see **Section 9.03.3.2**). Efficient percolation can occur at low pressures (<3 GPa) when the S and O contents of the metal are very high (i.e., close to the Fe–S eutectic), but such conditions are not applicable to core formation in a large planet such as the Earth. There is a preliminary indication that efficient percolation ($\theta < 60°$) may be a feasible mechanism under deep lower-mantle conditions, but this result awaits confirmation.

When dihedral angles significantly exceed 60°, experimental evidence suggests that liquid metal can separate from crystalline silicates when the material is undergoing shear deformation due to nonhydrostatic stress (Bruhn et al., 2000; Groebner and Kohlstedt, 2006; Hustoft and Kohlstedt, 2006; Rushmer et al., 2000). The most recent results indicate that ~1 vol.% liquid metal remains stranded in the silicate matrix so that reasonably efficient percolation of liquid metal, with percolation velocities on the order of 150 km year^{-1}, might occur in crystalline mantle that is undergoing solid-state convection (Hustoft and Kohlstedt, 2006). There

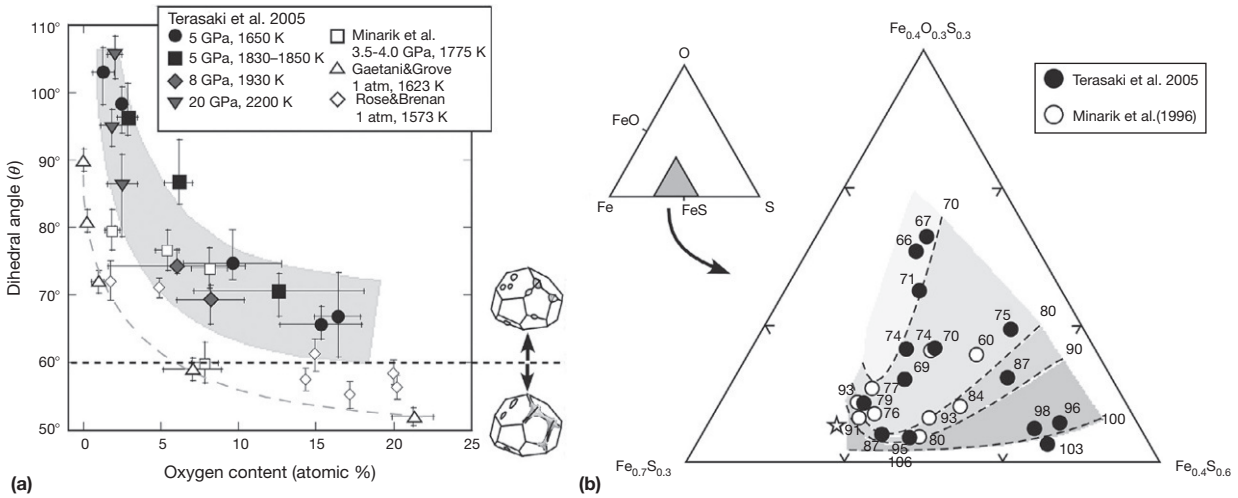


Figure 5 Dihedral angles in aggregates of olivine (≤8 GPa) and ringwoodite (20 GPa) containing several vol.% Fe–FeS melt as a function of the melt composition. (a) Dihedral angle as a function of the oxygen content of the Fe-alloy liquid. As the oxygen content increases, the structures of the silicate and liquid Fe alloy become more similar with the result that both the interfacial energy and the dihedral angle decrease. (b) Effects of O and S contents of the Fe-alloy liquid on dihedral angles. The data points are plotted on a triangular section of the Fe–O–S system, and each data point is labeled with the dihedral angle. The dashed lines are contours of constant dihedral angle. These results suggest that the effect of dissolved O on the dihedral angle is greater than that of dissolved S. Reproduced from Terasaki H, Frost DJ, Rubie DC, and Langenhorst F (2005) The effect of oxygen and sulphur on the dihedral angle between Fe–O–S melt and silicate minerals at high pressure: Implications for Martian core formation. *Earth and Planetary Science Letters* 232: 379–392, http://dx.doi.org/10.1016/j.epsl.2005.01.030.

are, however, two potential problems with this shear-induced percolation mechanism. First, experiments have been performed at high strain rates (10^{-2} to $10^{-5}\,s^{-1}$), and it is unlikely that the mechanism would also be effective at much lower strain rates. This is a consequence of deformation being strongly localized by the liquid phase at high strain rates, so that liquid interconnection and segregation occur, whereas at low strain rates, the geometry of liquid pockets is controlled predominantly by surface tension that results in little or no interconnection (Walte et al., 2011). Secondly, the conversion of potential energy to heat during percolation of iron liquid in a planet the size of Earth might be sufficient to melt the silicates and thus change the mechanism to the one discussed in the next section (see **Section 9.03.2.2.3**).

9.03.2.3.2 Metal-silicate separation in a magma ocean

According to the results of calculations of the energy released by giant impacts, it is clear that the collision of a Mars-sized body with the proto-Earth would have resulted in the melting of a large part of or even the entire planet (**Section 9.03.2.2.2**). In the case of partial melting, although the distribution of melt may initially have been concentrated on the side affected by the impact, isostatic readjustment would have led rapidly to the formation of a global magma ocean of approximately uniform depth (Tonks and Melosh, 1993). An important, and presently unresolved, question is whether this isostatic adjustment takes place on a timescale longer or shorter than that of iron separation. The mechanics of separation and chemical equilibration is quite different if the isostatic adjustment is much slower than iron separation, because then the melt region is not a global ocean of approximately uniform depth, but a restricted 'sea' that is both hotter and deeper than the ocean that eventually develops. This question is further raised in the succeeding text in **Section 9.03.3.2.3**.

Because of the large density difference between liquid iron and liquid silicate, magma ocean formation provides a rapid and efficient mechanism for the separation of these two phases. Here, we examine the physics of metal–silicate separation in some detail because of the consequences for interpreting mantle geochemistry, as discussed in the succeeding text in **Section 9.03.3.2**.

A fundamental property that controls the dynamic behavior of a deep magma ocean is the viscosity of ultramafic silicate liquid. The viscosity of peridotite liquid has been determined experimentally at 2043–2523 K and 2.8–13.0 GPa by Liebske et al. (2005). Their results (**Figure 6**) show that viscosity increases with pressure up to 9–10 GPa and then decreases to at least 13 GPa (Reid et al., 2003 found a similar trend for $CaMgSi_2O_6$ liquid). Based on a study of the self-diffusion of O and Si in silicate liquid, viscosity is expected to increase again at pressures above 18 GPa (Schmickler et al., 2005). The transient decrease in viscosity in the pressure range 9–18 GPa may be caused by pressure-induced coordination changes (e.g., formation of five- and sixfold coordinated Si) in the melt structure (Liebske et al., 2005). Unfortunately, it is currently not possible to extrapolate these experimental results reliably in order to make predictions of melt viscosities at pressures significantly higher than 18 GPa; the large uncertainties involved in doing this are illustrated by Liebske et al. (2005, their Figure 7). An alternative approach is to use first-principles

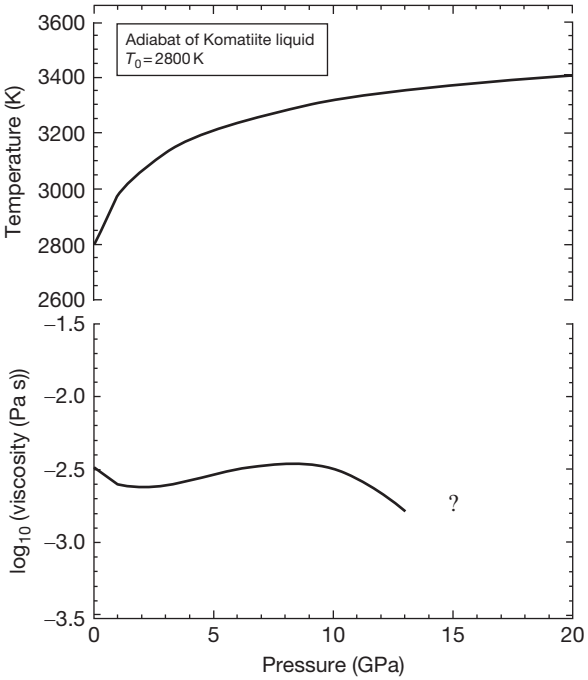


Figure 6 The viscosity of the upper part of a peridotitic magma ocean that has a total depth of ~1800 km (bottom). The viscosity is shown only to a depth of ~400 km because the existing experimental data cannot be extrapolated reliably to higher pressures. Based on the results of Schmickler et al. (2005), the viscosity likely increases again above ~18 GPa, as suggested also by molecular dynamics simulations (Karki and Stixrude, 2010). The viscosity profile is based on the adiabat shown in the top part of the figure. '?' means that the continuation of the lower curve to higher pressures is unknown. Reproduced from Liebske C, Schmickler B, Terasai H, et al. (2005) Viscosity of peridotite liquid up to 13 GPa: Implications for magma ocean viscosities. *Earth and Planetary Science Letters* 240: 589–604.

molecular dynamics simulations to estimate melt viscosity at conditions beyond those achievable experimentally. The viscosity of $MgSiO_3$ liquid has thus been determined by Karki and Stixrude (2010) to 150 GPa and was found to vary by two orders of magnitude over the pressure range of the Earth's mantle.

On the basis of the experimental data of Liebske et al. (2005), the viscosity of a magma ocean, at least to a depth of ~500 km, is estimated to lie in the range 0.01–0.003 Pa s, which is extremely low (for comparison, the viscosity of water at ambient conditions is 0.001 Pa s). Consequently, the Rayleigh number (which provides a measure of the vigor of convection) is extraordinarily high, on the order of 10^{27}–10^{32} (e.g., Rubie et al., 2003; Solomatov, 2000; **Chapter 9.04**). This means that a deep magma ocean undergoes vigorous turbulent convection, with convection velocities on the order of at least a few meters per second. Using a simple parameterized convection model, the rate of heat loss can also be estimated, which leads to the conclusion that, in the absence of other effects (see succeeding text), the lifetime of a deep magma ocean on Earth is only a few thousand years (Solomatov, 2000).

What is the physical state of molten iron in a vigorously convecting magma ocean? Initially, iron metal may be present in states that range from finely dispersed submillimeter

particles (as in undifferentiated chondritic material) to large molten masses that originated as cores of previously differentiated bodies (ranging in size from planetesimals to Mars-sized planets) that impacted the accreting Earth. In a molten system, very small particles tend to grow in size by coalescing with each other in order to reduce surface energy. Large molten bodies, on the other hand, are unstable as they settle and tend to break up into smaller bodies. A crucial question concerns the extent to which an impactor's core breaks up and becomes emulsified as it travels through the target's molten mantle (see Karato and Murthy, 1997; Rubie et al., 2003; Stevenson, 1990, for discussions of this issue). Hallworth et al. (1993) noted that laboratory-scale turbidity currents travel only a few times their initial dimension before being dispersed by turbulent instabilities. Such a core will experience both shear (Kelvin–Helmholtz) and buoyancy (Rayleigh–Taylor) instabilities. These processes operate at different length scales (R–T instabilities are small-scale features; see Dalziel et al., 1999), but both processes will tend to break the body up until a stable droplet size is reached at which surface tension inhibits further breakup. The stable droplet size can be predicted using the dimensionless Weber number, which is defined as

$$W_e = \frac{(\rho_m - \rho_s) d v_s^2}{\sigma} \quad [6]$$

where ρ_m and ρ_s are the densities of metal and silicate, respectively, d is the diameter of metal bodies, v_s is the settling velocity, and σ is the surface energy of the metal–silicate interface (Young, 1965). A balance between coalescence and breakup is reached when the value of W_e is approximately 10: When the value is larger than 10, instabilities cause further breakup to occur, and when it is less than 10, coalescence occurs. The settling velocity v_s is determined using Stokes' law when the flow regime is lamellar or an equation that incorporates a drag coefficient when the flow around the falling droplet is turbulent (Rubie et al., 2003). Both the settling velocity and the droplet size depend on silicate melt viscosity. For likely magma ocean viscosities and assuming a metal–silicate surface energy of 1 $N\,m^{-1}$, the droplet diameter is estimated to be ~1 cm and the settling velocity ~0.5 $m\,s^{-1}$ (**Figure 7**; Rubie et al., 2003, see also Karato and Murthy, 1997; Stevenson, 1990). However, exact value of the surface energy in eqn [6] is dependent on the concentrations of elements such as S, Si, and O in the liquid metal, as shown experimentally up to 1.5 GPa (Terasaki et al., 2009, 2012). Thus, settling velocities and droplet sizes could deviate from the aforementioned values by a factor of 2, depending on the liquid metal composition (Terasaki et al., 2012).

Having estimated the stable size of metal droplets, the next question is as follows: how quickly does a large mass of metal (e.g., 50–500 km in diameter) become emulsified and break up into a 'rain' of small droplets of stable diameter? Although the process is currently not well understood, Rubie et al. (2003) argued that emulsification should occur within a falling distance equal to a few times the original diameter of the body. Thus, the cores of all, except perhaps the largest, impacting bodies should have experienced a very large degree of emulsification.

Several authors have recently approached the question of emulsification of large bodies of iron sinking in a silicate magma ocean. Dahl and Stevenson (2010) argued that emulsification during accretion is incomplete, with only 1–20% of the Earth's core having equilibrated with the mantle. They applied analytic fluid dynamic models to investigate the dispersion of masses of iron originating from the cores of large accreting planetesimals. They analyzed both the Rayleigh–Taylor and the Kelvin–Helmholtz instabilities to estimate the mixing of an initially compact spherical blob of iron in a

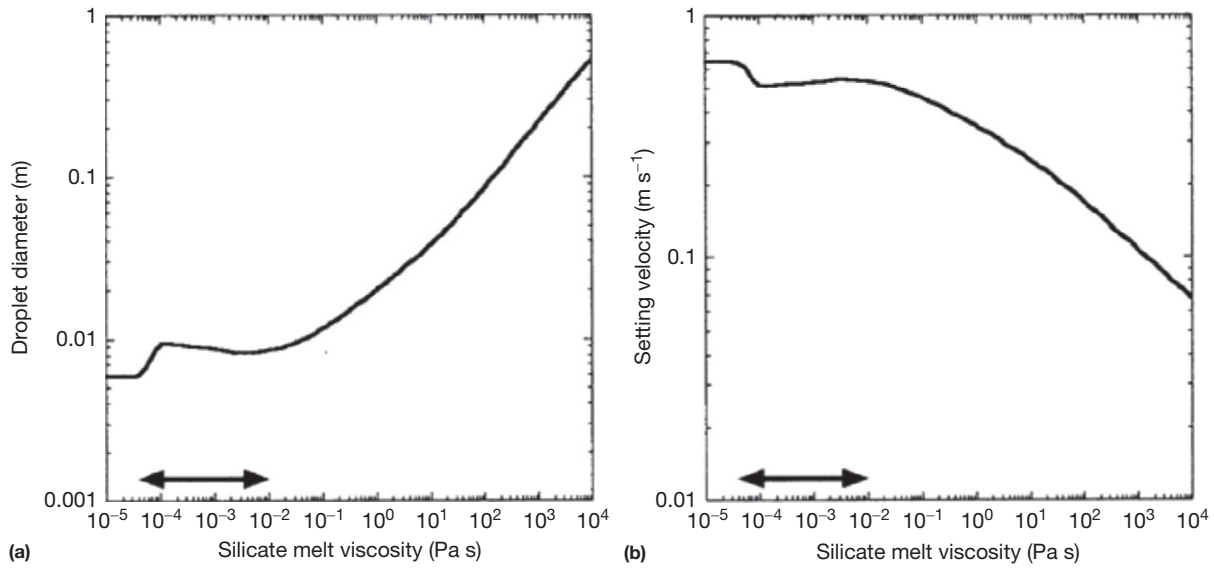


Figure 7 (a) Stable diameter of liquid iron droplets dispersed in a magma ocean as a function of silicate melt viscosity. The droplet diameter is calculated using the Weber number, as explained in the text. (b) Terminal settling velocity of liquid Fe droplets (of stable droplet diameter) as a function of silicate melt viscosity. The nonlinear trend arises from a transition from Stokes to turbulent flow at low viscosities. In both (a) and (b), the arrowed lines indicate the range of likely magma ocean viscosities. Reproduced from Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003). Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

silicate liquid magma ocean and concluded that the blob must sink many times its own diameter before it is reduced to droplets small enough for chemical equilibration. On this basis, they argued that planetesimal cores larger than about 10 km diameter are unable to chemically equilibrate with the surrounding silicate. Their analytic models were qualitatively confirmed in a suite of experiments by Deguen et al. (2011) who created laboratory-scale models of the turbulent descent of an initially concentrated mass of dense, immiscible liquid into a tank of less dense fluid. In contrast, Samuel (2012) numerically modeled the breakup of iron diapirs and argued instead that initially concentrated iron masses smaller than the depth of the magma ocean may rapidly break up and emulsify during sinking under plausible conditions. None of these models, however, properly address the initial conditions following the high-speed impact of a planetesimal with a pre-differentiated core upon the accreting Earth. Preliminary work by Kendall and Melosh (2012a,b) examined the dispersion of planetesimal cores of various sizes during accretion using both vertical 2D and oblique 3D numerical impact models. They concluded that the cores are stretched by factors of about 1000 during such impacts and that the proper initial condition is that of an already-mixed cloud of liquid metal blobs dispersed in the silicate magma ocean. In this case, the impact itself produces a large reduction of the initial size of the planetesimal cores near the top of the magma ocean. Subsequent dispersion during sinking may then reduce the size of the iron masses to the point where chemical equilibration will readily occur. It is clear that understanding this crucial question is still in a state of flux and further progress will be needed to settle the issue of emulsification of the cores of impacting objects.

The primary importance of emulsification is that it determines the degree to which chemical and thermal reequilibration occurs (Karato and Murthy, 1997; Rubie et al., 2003). Because thermal diffusivities are higher than chemical diffusivities, thermal equilibrium is always reached first. For typical settling velocities, iron blobs on the order of ~0.01 m in diameter will remain in chemical equilibration with the surrounding silicate liquid as they fall (Rubie et al., 2003). Droplets that are much larger (e.g., on the order of meters or more in diameter) will fall rapidly and will not remain in equilibrium with the adjacent silicate because diffusion distances are too large. The physical arguments for emulsification summarized here are supported by evidence for chemical equilibration from both Hf–W observations (**Section 9.03.3.1**) and siderophile element abundances in the mantle (**Section 9.03.3.2**). Since chemical equilibration of macroscale iron bodies is very slow, this apparent equilibration strongly suggests that the iron was present, at least to a large extent, as small dispersed droplets (Righter and Drake, 2003).

Because likely settling velocities (~0.5 $m\,s^{-1}$) of small iron droplets are much lower than typical convection velocities (~10 $m\,s^{-1}$), iron droplets may remained entrained for a significant time in the magma ocean, and accumulation through sedimentation at the base of the ocean will be a slow and gradual process. The dynamics of the settling and accumulation processes are important because they determine the chemical consequences of core formation (e.g., siderophile element geochemistry), as discussed in the succeeding text in **Section 9.03.3.2.3**.

The time taken for a magma ocean to start to crystallize is an important parameter when evaluating metal–silicate equilibrium models, as discussed in the succeeding text (**Section 9.03.3.2.3**). The existence of an early, thick atmosphere has little effect on large incoming impactors but may be sufficiently insulating that, by itself, it ensures a magma ocean (e.g., Matsui and Abe, 1986). The depth to the (rheologically determined) base of the magma ocean is determined by the point at which the adiabat crosses the geotherm defining a melt fraction of roughly 60% (Solomatov, 2000). The survival time of the magma ocean depends on both the atmosphere and whether or not an insulating lid can develop. In the absence of these two effects, the lifetimes are very short, on the order of 10^3 years (e.g., Pritchard and Stevenson, 2000; Solomatov, 2000). However, if a conductive lid develops, the lifetime may be much longer, on the order of 10^8 years (Spohn and Schubert, 1991), and similar lifetimes can arise due to a thick atmosphere (Abe, 1997). Thus, the lifetime of magma oceans is currently very unclear. The Moon evidently developed a chemically buoyant, insulating crust on top of its magma ocean (Warren, 1985). However, it did so because at low pressures aluminum partitions into low-density phases, especially plagioclase. At higher pressures, Al instead partitions into dense garnet, in which case a chemically buoyant crust will not develop (e.g., Elkins-Tanton et al., 2003, 2011). In the absence of chemical buoyancy, a solid crust will still develop but will be vulnerable to disruption by impacts or foundering (Stevenson, 1989). The latter process in particular is currently very poorly understood, and thus, the lifetime of magma oceans remains an open question. Fortunately, even the short-lived magma oceans persist for timescales long compared to most other processes of interest.

The conventional view has been that a deep magma ocean would crystallize from the bottom up, as suggested by the relationship between the adiabat and melting curves shown in **Figure 8(b)**. However, recent investigations of the thermodynamic properties of Mg_2SiO_4 and $MgSiO_3$ liquids at high pressure, through shock experiments (Mosenfelder et al., 2007, 2009) and molecular dynamics simulations (Stixrude and Karki, 2005), suggest that temperatures along a magma ocean adiabat increase much more rapidly than shown in **Figure 8(b)**. These results lead to the possibility that a magma ocean starts to crystallize at mid-depths. This proposal has led to the idea that a basal magma ocean existed above the CMB early in Earth's history (Labrosse et al., 2007).

9.03.2.3.3 Diapirs and diking

If percolation is not an effective mechanism, then differentiation may occur either by downward migration of large iron blobs (diapirism) or by propagation of iron-filled fractures (diking). The mode of descent of a dense, low-viscosity fluid through a less dense high-viscosity fluid has been the subject of much debate. Many authors have assumed that the dense fluid (liquid iron in our case) penetrates the underlying less dense material (unmelted mantle) as diapirs, somewhat analogous to the (inverted) penetration of high-viscosity sediments by low-density salt diapirs or the penetration of the lower crust by granitic magma above subduction zones. Rubin (1993), however, argued that the mode of penetration, dikes vs. diapirs, depends mainly upon the ratio of viscosity between the intruding fluid and the host material.

The transport of iron though crystalline mantle as large diapirs, 1–10 km in diameter or larger, has been discussed in

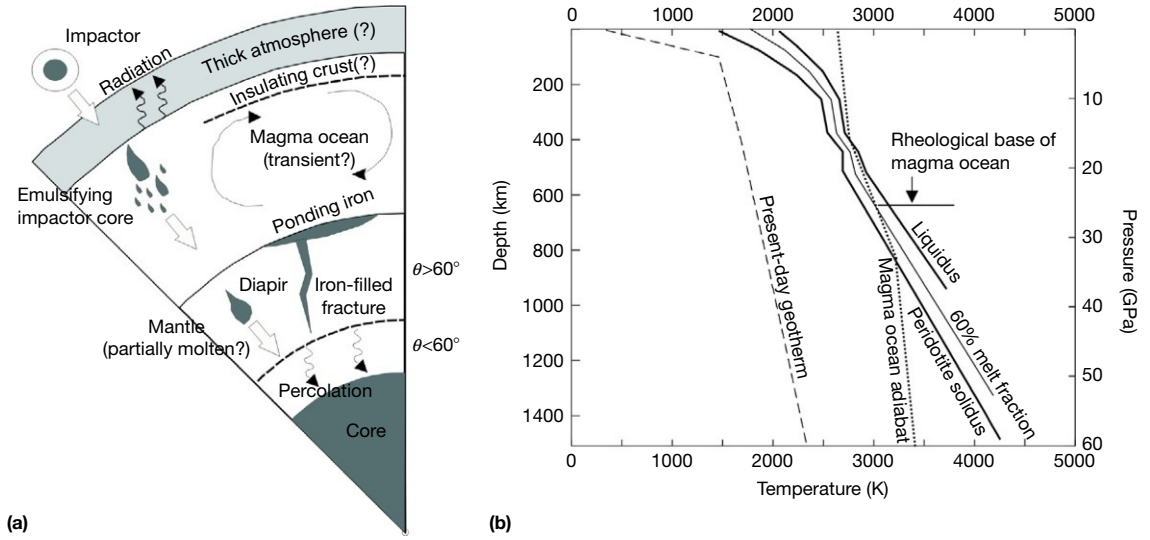


Figure 8 (a) Schematic diagram showing the various processes of metal–silicate segregation that could have operated during Earth accretion and core formation. The rheological base of the magma ocean is defined as the point at which the melt fraction drops below ~60% (Solomatov, 2000). (b) Schematic illustration of the temperature structure in the early Earth, adapted from Walter MJ and Tronnes RG (2004) Early Earth differentiation. *Earth and Planetary Science Letters* 225: 253–269. The solidus curve in the lower mantle is based on Boehler (2000), but according to more recent determinations (Fiquet et al., 2010; Andrault et al., 2011; Liebske and Frost, 2012), solidus temperatures are considerably lower (e.g., ~3200 K at 60 GPa). The 60% melt fraction line is schematic.

detail by Karato and Murthy (1997). When liquid iron ponds as a layer at the base of a magma ocean (Figure 8), gravitational instabilities develop due to the density contrast with the underlying silicate-rich material and cause diapir formation. Their size and rate of descent through the mantle depend on the initial thickness of the metal layer and the viscosity of the silicate mantle. Clearly, gravitational heating will be important and will facilitate diapir descent by reducing the viscosity of the adjacent mantle. In contrast to magma ocean segregation, there will be no significant chemical exchange between metal and silicate, chemical disequilibrium will result, and siderophile element abundances in the mantle cannot be a consequence of this mechanism (Karato and Murthy, 1997; see also Chapter 9.02).

Liquid iron ponded at the base of the magma ocean may also, under the right conditions, sink rapidly toward the Earth's core by diking. Although it may be supposed that the hot, but nevertheless crystalline, mantle underlying the magma ocean cannot support brittle cracks, numerical studies summarized in Rubin (1995) indicate that dikes can still form, so long as the contrast in viscosity between the fluid in the dike and the surrounding host rocks is greater than 10^{11}–10^{14}. With a viscosity around 10^{-2} Pa s, liquid iron is thus expected to form dikes if the viscosity of the host rock exceeds 10^{9}–10^{12} Pa s. Given that the viscosity of the asthenosphere today is around 10^{19} Pa s, it is not unreasonable to expect the iron to reach the core via narrow dikes rather than as diapirs. In this case, even less time is required for the rapidly descending iron to reach the core and thus less time for the iron to chemically equilibrate with the surrounding mantle. Indeed, even in the present Earth, Stevenson (2003) has proposed that masses of molten iron as small as 10^{8} kg (which would fill a cube about 25 m on a side) could travel from the Earth's surface to the core in about 1 week.

9.03.2.3.4 Summary and implications for chemical equilibration

A schematic illustration of how the various differentiation mechanisms might operate together is shown in Figure 8. Liquid metal separates rapidly from liquid silicate in a deep magma ocean and accumulates as ponded layers at the rheological base of the magma ocean. The ponded iron then migrates through the largely crystalline underlying mantle toward the protocore by either percolation, diapirism, or diking. According to experimental results summarized earlier, percolation is unlikely to be a completely efficient mechanism, even when catalyzed by shear deformation, but the mechanism may at least contribute to core formation. Whether the diking mechanism shown in Figure 8 is dominant or the iron descends in diapirs initiated by Rayleigh–Taylor instabilities depends upon poorly known material properties, such as viscosity, that make it difficult to make definite inferences about the mechanism. What seems to be clear is that, once liquid iron accumulates in large masses on the floor of the magma ocean, it then descends rapidly to the Earth's core without further chemical interaction with the underlying silicate.

Despite the rapid transit times associated with falling iron drops and the somewhat longer timescales of percolative flow, the inferred length scales are small enough that complete chemical equilibration is expected. Conversely, descending iron diapirs are sufficiently large that chemical equilibrium is expected to be negligible (Karato and Murthy, 1997). Thus, the different differentiation mechanisms have very different chemical consequences. However, the magnitude of these chemical effects also depends on the relative abundances: a late passage of 1% core material through the mantle may well have a strong effect on mantle siderophile element abundances (e.g., W or Pt) but will have little effect on major element concentrations (e.g., oxygen) simply because the core material will become

saturated and thus transport insignificant amounts of these more abundant elements.

9.03.3 Observational and Experimental Constraints

Having discussed our theoretical expectations of accretion and core formation processes, we will now go on to discuss the extent to which observations and experimental data may be used to differentiate between the various theoretical possibilities. Excellent summaries of many of these observations and experimental results may be found in Halliday (2003) and in the volume edited by Canup and Righter (2000).

9.03.3.1 Core Formation Timescales

An extremely important development has been the recognition that some isotopic systems provide an observational constraint on core formation timescales and thus planetary accretion rates. The most useful isotopic system is Hf–W (Jacobsen, 2005; Kleine et al., 2009). The Pd–Ag system is similar (Schönbächler et al., 2010) but is more complicated to interpret because silver is a volatile element. The U–Pb system has similar volatility problems but generates results that can be reconciled with the more robust Hf–W technique (Rudge et al., 2010; Wood and Halliday, 2005, 2010).

The Hf–W chronometer works as follows. W is siderophile (i.e., 'metal-loving'), while Hf is lithophile (it remains in silicates). Furthermore, ^{182}Hf decays to stable ^{182}W with a half-life of 9 Myr. If an initially undifferentiated object suddenly forms a core after all the ^{182}Hf has decayed, the W will be extracted into the core and the mantle will be strongly depleted in all tungsten isotopes. However, if core formation occurs early, while ^{182}Hf is live, then the subsequent decay of ^{182}Hf to ^{182}W will enrich the mantle in radiogenic tungsten compared with nonradiogenic tungsten. Thus, a radiogenic tungsten excess, or tungsten anomaly, in the mantle is a sign of early core formation. Furthermore, if the silicate–iron mass ratio and the mantle concentrations of Hf and W compared with undifferentiated materials (chondrites) are known, then the observed tungsten anomaly can be used to infer a single-stage core formation age. In the case of the Earth, this single-stage age is roughly 30 Myr (Jacobsen, 2005). The lunar Hf–W record is difficult to interpret but may imply a formation time for the Moon of 50–150 Myr after solar system formation (Touboul et al., 2007).

There are three characteristics of the Hf–W system that makes it especially suitable for examining core formation. Firstly, the half-life is comparable to the timescale over which planets are expected to form. Secondly, there are few other processes likely to lead to tungsten fractionation and perturbation of the isotopic system, though very early crustal formation or neutron capture (Kleine et al., 2005) can have effects. Finally, both Hf and W are refractory; certain other isotopic systems suffer from the fact that one or more elements (e.g., lead and silver) are volatile and can be easily lost during accretion.

The fact that tungsten isotope anomalies exist in the terrestrial mantle implies that core formation was essentially complete before about five half-lives (50 Myr) had elapsed. Mars and Vesta have larger tungsten anomalies, indicating that core formation ended earlier on these smaller bodies (Dauphas and Pourmand, 2011; Kleine et al., 2009). The timescales implied are compatible with the theoretical picture of planetary accretion described in **Section 9.03.2.1**. The simple model of a single core formation event is of course a simplification of the real picture, in which the bulk of the core mass is added during stochastic giant impacts. However, more complicated models, in which the mass is added in a series of discrete events, do not substantially alter the overall timescale derived.

The observed tungsten anomaly depends mainly on the timescale over which the core forms, the relative affinities of Hf and W for silicates, and the extent to which the cores of the impactors reequilibrate with the target mantle. The relative affinities of Hf and W can be determined, in a time-averaged sense, by measuring the present-day concentrations of these elements in the mantle. These affinities (i.e., the partition coefficients) may have varied with time, due to changing conditions (P, T, and oxygen fugacity f_{O_2}) in the Earth. Although the dependence of the partition coefficients on these variables is known (e.g., Cottrell et al., 2009; Wade et al., 2012), how the variables actually evolved as the Earth grew is very poorly understood (e.g., Halliday, 2004). This caveat aside, if one accepts that the accretion timescales determined by numerical accretion models are reasonable, these models can then be used to investigate the extent to which reequilibration must have occurred.

Figure 9 shows examples of the tungsten anomalies generated from numerical models of late-stage accretion (Nimmo et al., 2010). In **Figure 9(a)**, it is assumed that 50% of incoming core material reequilibrates with the silicate mantle. Because of the stochastic nature of the accretion process, there is considerable scatter; nonetheless, bodies that resemble Mars and Earth in both their tungsten anomalies and the Hf/W ratios can be produced. **Figure 9(b)** shows how the average tungsten anomaly for Earth-mass bodies varies as the equilibration factor k is varied. If little equilibration occurs, the resulting tungsten anomalies are large. This is because the larger bodies then inherit their anomalies from smaller bodies that (by assumption) differentiated early and thus developed large tungsten anomalies. By contrast, if equilibration is efficient, then the isotopic memory of earlier events is erased and each impact tends to drive down the overall tungsten anomaly. Assuming that the accretion timescales generated by the simulations are correct, **Figure 9(b)** suggests that an intermediate reequilibration amount ($k \approx 50\%$) is required to reproduce the observations; similar results have been obtained by other authors (e.g., Rudge et al., 2010). This conclusion in turn places constraints on the physics of these very large impacts and, in particular, suggests that some emulsification of the impacting core occurs as it travels through the magma ocean (see also Rubie et al., 2011; **Section 9.03.2.3.2**).

One example of the importance of the equilibration factor k is in the accretion of Mars. Mars has an Earthlike tungsten anomaly but a lower Hf/W ratio, implying rapid accretion. If equilibration is efficient ($k \approx 1$), then accretion must have been so rapid that Mars would have melted via ^{26}Al decay – this is the hypothesis advanced by Dauphas and Pourmand (2011). However, less extreme reequilibration would permit more prolonged accretion and thus reduce the likelihood of melting. A better understanding of k is thus crucial to understanding Mars' early evolution.

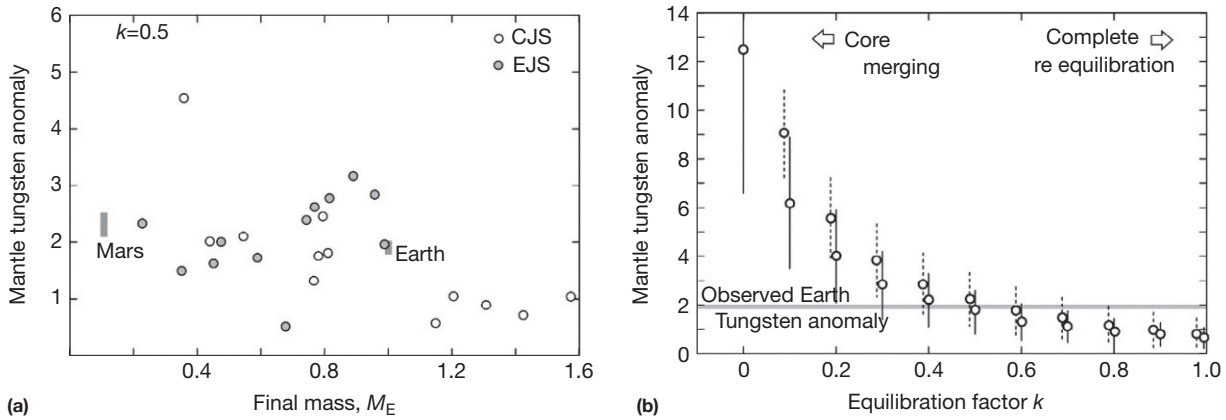


Figure 9 Isotopic outcomes of core formation based on N-body accretion codes (from Nimmo et al., 2010). M_E is the mass of the final bodies in Earth masses; k is the equilibration factor between core and mantle (see text). Shaded regions are observed values; circles are model results. (a) Individual outcomes with $k = 0.5$. Open and filled circles are based on N-body simulations with Jupiter and Saturn on their current orbits and on smaller, eccentric orbits, respectively (see O'Brien et al., 2006). (b) Mean and standard deviations of tungsten anomalies for bodies with masses >0.7 M_E as a function of equilibration factor. Solid and dashed lines are for models with a variable and fixed tungsten partition coefficient, respectively (see Nimmo et al., 2010).

The palladium–silver (^{107}Pd–^{107}Ag) system is in some ways very similar to the Hf–W system. Pd is strongly siderophile and ^{107}Pd has a half-life of 6.5 Myr, making it suitable for investigating core formation processes. However, there are fewer isotopes of Ag than of W, making precise isotopic measurements more difficult; more importantly, Ag is volatile which complicates interpretation of the isotopic data. Nonetheless, precise ^{107}Ag/^{109}Ag measurements have been used to argue for an Earth accretion timescale compatible with that derived from Hf–W systematics and also involving a change from volatile-depleted to volatile-enriched material as accretion proceeded (Schönbächler et al., 2010). This latter conclusion is entirely consistent with dynamical simulations in which the Earth's feeding zone expanded outward beyond the 'snow line' with time (O'Brien et al., 2006).

Finally, because the early history of the core and the mantle is intimately coupled, there are some isotopic systems governed by mantle processes that are also relevant to core formation timescales. In particular, the nonchondritic ^{142}Nd isotope signature of the Earth's upper mantle has been used to argue for global melting of the mantle within 30 Myr of solar system formation (Boyet and Carlson, 2005). Sm–Nd and Lu–Hf chronometers are also consistent with the solidification of a magma ocean within the first ~100 Myr of Earth history (Caro et al., 2005), and U–Pb dates have been interpreted as resulting from the final stage of magma ocean crystallization at about 80 Myr after solar system formation (Wood and Halliday, 2005, 2010). Xe isotope data give a comparable time for loss of xenon from the mantle (e.g., Halliday, 2003; Porcelli et al., 2001) and also support early differentiation scenarios (Mukhopadhyay, 2012) and multiple magma ocean generations (Tucker and Mukhopadhyay, 2014).

In summary, the Hf–W (and, to a lesser extent, Pd–Ag and U–Pb) systems are important for two reasons: they constrain the timescale over which core formation occurred and also constrain the extent of reequilibration between core material and mantle material. To obtain the pressure–temperature conditions under which this reequilibration took place, it is necessary to look at other siderophile elements as well. Inferring the equilibration conditions is important both because constraints are then placed on the early thermal state of both core and mantle and because these conditions strongly influence the ultimate composition of the core.

9.03.3.2 Constraints from Siderophile Element Geochemistry

9.03.3.2.1 Introduction to siderophile element geochemistry

The primary geochemical evidence for core formation in the Earth is provided by a comparison of the composition of the Earth's silicate mantle with its bulk composition. Estimates of the composition of the mantle are based on numerous geochemical studies of mantle peridotites and are well established (e.g., McDonough and Sun, 1995; Palme and O'Neill, 2003), assuming of course that the mantle is homogeneous. Although the bulk composition of the Earth is not known precisely (Drake and Righter, 2002; O'Neill and Palme, 1998), it is often approximated by the refractory-element composition of C1 carbonaceous chondrites, which are the most pristine (undifferentiated) relicts known from the early solar system. As seen in **Figure 10**, refractory lithophile elements (e.g., Al, Ca, Ti, Ta, Zr, and the rare earth elements, REEs) are present in the mantle in CI chondritic concentrations, and their concentrations have therefore been unaffected by accretion or differentiation processes. Compared with the bulk composition of the Earth, the mantle is strongly depleted in (a) siderophile (metal-loving) elements that have partitioned into iron-rich metal during formation of the Earth's core (Walter et al., 2000) and (b) volatile elements that are considered to have been partly lost during accretion. Note that some of the volatile elements are also siderophile (e.g., sulfur) so that current mantle concentrations can be the result of both core formation and volatility (for a full classification of depleted elements, see Table 1 in Walter et al. (2000)). Siderophile elements that are unaffected by volatility are most valuable for understanding core formation, and these include the moderately siderophile elements (MSEs; e.g., Fe, Ni, Co, W, and Mo) and the HSEs, which include the platinum group elements (e.g., Re, Ru, Rh, Os, Ir, Pd, Pt, and Au).

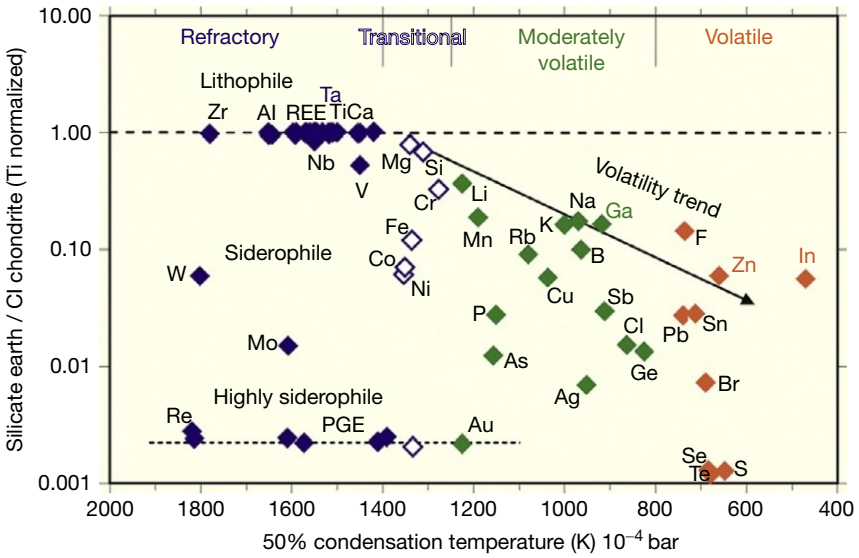


Figure 10 Element abundances of the Earth's mantle normalized to CI chondrite and the refractory element Ti (data after Palme and O'Neill, 2003) plotted against their 50% condensation temperatures as given by Wasson (1985) and, in the case of Pb, by Lodders (2003). Note that the errors on these data points may be large (e.g., at least an order of magnitude for some volatile elements). Courtesy of Ute Mann.

The degree of siderophile behavior is described for element M by the metal–silicate partition coefficient $D_M^{met-sil}$, which is defined as

$$D_M^{met-sil} = \frac{C_M^{met}}{C_M^{sil}} \quad [7]$$

where C_M^{met} and C_M^{sil} are the wt% or molar concentrations of M in metal and silicate, respectively. MSEs are defined as having values (determined experimentally at 1 bar) of $D_M^{met-sil} < 10^4$, whereas HSEs have 1 bar values that are greater than 10^4 and can be, in the case of Ir, for example, as high as 10^{10}. The boundary between siderophile behavior and lithophile behavior is defined as $D_M^{met-sil} = 1$.

As discussed in the succeeding text, partition coefficients are a function of pressure, temperature, and the compositions of the metal and silicate phases. An additional controlling parameter, which is critical in the context of core formation, is oxygen fugacity, f_{O_2}. The effect of f_{O_2} on the partitioning of a siderophile element depends on its valence state when it is dissolved in the silicate phase, as shown by the following metal–silicate equilibrium:

$$\underset{\text{metal}}{M} + (n/4)O_2 = \underset{\text{silicate liquid}}{MO_{n/2}}$$

(e.g., Righter and Drake, 2003). Here, M is the element of interest and n is its valence when it is dissolved in silicate liquid. Increasing the oxygen fugacity drives the equilibrium toward the right-hand side, thus increasing the concentration of M in the silicate liquid and reducing the value of $D_M^{met-sil}$. Typically, the dependence of $D_M^{met-sil}$ on f_{O_2}, pressure (P), and temperature (T) is expressed by a relationship of the form

$$\ln D_M^{met-sil} = a \ln fO_2 + b/T + cP/T + g \quad [8]$$

where a is related to the valence state of M in the silicate liquid and b, c, and g are related to thermodynamic free energy terms that have generally assumed to be constants even when the P–T range of extrapolation is large (see Righter and Drake, 2003). In reality, these parameters are not likely to be constant over large ranges of pressure, temperature, and f_{O_2}. For example, the valence state of an element can change as a function of f_{O_2}. Pressure-induced changes in the structure of the silicate melt and/or the metal liquid may also cause a strong nonlinear pressure dependence at certain conditions (Kegler et al., 2008; Keppler and Rubie, 1993; Sanloup et al., 2011). In addition to P, T, and f_{O_2}, the effects of additional factors, such as the composition/structure of the silicate melt and the sulfur, silicon, and carbon contents of the metal, may also need to be included (Cottrell et al., 2009; Tuff et al., 2011; Wade et al., 2012; Walter et al., 2000).

For core formation, the value of f_{O_2} is constrained by the partitioning of Fe between the mantle and the core (see Mann et al., 2009) and has been estimated to have been 1–2 log units below the oxygen fugacity defined by the iron–wüstite (Fe–FeO or 'IW') buffer (which is abbreviated as IW-1 to IW-2). However, in recent models, f_{O_2} is considered to have been initially very low and to have increased during core formation to a final value of around IW-2 (see succeeding text).

The oxygen fugacity term in eqn [8] can be eliminated by considering the exchange reaction

$$\underset{\text{silicate}}{MO_{n/2}} + \frac{n}{2}\underset{\text{metal}}{Fe} = \underset{\text{metal}}{M} + \frac{n}{2}\underset{\text{silicate}}{FeO} \quad [9]$$

and by describing the partitioning of element M in terms of a distribution coefficient K_D, instead of $D_M^{met-sil}$, where

$$K_D = \frac{X_M^{met}\left(X_{FeO}^{sil}\right)^{n/2}}{X_{MO_{n/2}}^{sil}\left(X_{Fe}^{met}\right)^{n/2}} \quad [10]$$

(e.g., Mann et al., 2009). Here, X represents the mole fractions of M, $MO_{n/2}$, Fe, and FeO in metal (met) and silicate (sil). In a system in which mixing is ideal or the ratios of activity coefficients are constant, K_D is independent of composition and

oxygen fugacity and its pressure and temperature dependences can be described by an expression similar to eqn [8] but without the oxygen fugacity term (e.g., Mann et al., 2009). However, when the activities of components in the silicate and metal are dependent on composition, an equilibrium constant K must be used and is defined as

$$K = \frac{X_{M}^{met}\gamma_{M}^{met}\left(X_{FeO}^{sil}\gamma_{FeO}^{sil}\right)^{n/2}}{X_{MO_{n/2}}^{sil}\gamma_{MO_{n/2}}^{sil}\left(X_{Fe}^{met}\gamma_{Fe}^{met}\right)^{n/2}} \quad [11]$$

where γ represents the activities of the components in silicate and metal. Activity coefficients for the metal phase are formulated using the epsilon model (Wagner, 1962) as modified by Ma (2001) (e.g., Siebert et al., 2011; Tuff et al., 2011; Wade et al., 2012). For Mo and W, the silicate melt composition has a strong influence on partitioning and activity models have been discussed by Wade et al. (2012).

Note that carbon dissolved in liquid metal has a very large effect on the metal–silicate partitioning of Si and many siderophile elements. Consequently, experimental studies that have utilized graphite capsules can produce erroneous results (see Cottrell et al., 2009; Mann et al., 2009).

Values of $D_{M}^{met-sil}$ for the core–mantle system lie in the range 13–30 for MSEs and 600–1000 for HSEs (for compilations of values, see Wade and Wood, 2005; Wood et al., 2006). These values are much lower than experimentally determined 1 bar metal–silicate partition coefficients, in some cases by a few orders of magnitude, and show that the mantle contains an apparent *overabundance* of siderophile elements. Possible explanations of this anomaly, the so-called excess siderophile element problem, form the basis of theories of how and under which conditions the Earth's core formed (e.g., Newsom, 1990; Righter, 2003), as described further in **Section 9.03.3.2.2**. An important additional constraint is provided by the observation that the HSEs are present in the mantle in approximately chondritic relative abundances (**Figures 10** and **11**), even though the experimentally determined 1 bar $D_{M}^{met-sil}$ values vary by orders of magnitude.

Important constraints on how the Earth's core formed are potentially provided by also considering the geochemical consequences of core formation in other planetary bodies. If the sizes of such bodies differ significantly from that of the Earth, pressure–temperature conditions of metal–silicate equilibration will also be different, thus enabling some of the theories detailed in the succeeding text (**Section 9.03.3.2.2**) to be tested. Estimates of the geochemical effects of core formation in Mars, Vesta, and the Moon are also based on comparisons of siderophile element concentrations in the mantles of these bodies with their likely bulk compositions (e.g., Drake, 2001; Righter and Drake, 1996; Treiman et al., 1987; Walter et al., 2000). Both bulk and mantle compositions are poorly constrained compared with the Earth. Mantle compositions are inferred from studies of SNC meteorites for Mars, the eucritic meteorites for Vesta, and samples collected from the Moon's surface (McSween, 1999; Warren, 1993). In all cases, the samples (e.g., lavas and cumulates) are the products of crustal differentiation processes, and therefore, mantle compositions have to be inferred by taking the effects of differentiation into account. In the case of the Moon, there is an additional problem because the petrogenesis of samples is poorly constrained because of their extremely small size (Warren, 1993).

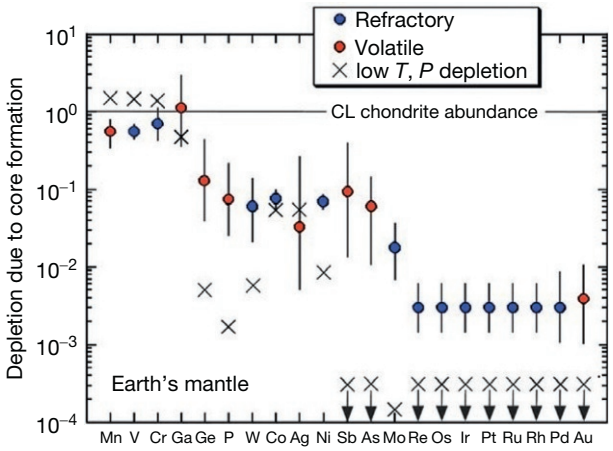


Figure 11 Depletion of siderophile elements in the Earth's mantle as the result of core formation. The siderophile elements are distinguished as being either refractory or volatile. The crosses show depletion values that would be expected on the basis of partitioning experiments performed at one bar and moderate temperatures (e.g., 1300 °C); in the case of the highly siderophile elements, the predicted values plot far below the bottom of the graph and are highly variable. The large discrepancies between the observed and the calculated values shown here form the basis for the 'siderophile element anomaly.' For further details, see Walter et al. (2000). Reproduced from Walter MJ, Newsom HE, Ertel W, and Holzheid A (2000). Siderophile elements in the Earth and Moon: Metal/silicate partitioning and implications for core formation. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 265–290. Tucson: University of Arizona Press; courtesy of Michael Walter.

Compared to the Earth's mantle, current assessments indicate that the Martian mantle is relatively depleted in Ni and Co, whereas concentrations of HSEs are similar. As in the Earth's mantle, the HSEs appear to be present in chondritic relative abundances (Jones et al., 2003; Kong et al., 1999; McSween, 1999; Warren et al., 1999). The mantles of both the Moon and Vesta show relatively large depletions in most siderophile elements (Righter and Drake, 1996; Walter et al., 2000). It is not possible in any of these cases to explain mantle siderophile element abundances by simple metal–silicate equilibrium at moderate pressures and temperatures, that is, based on 1 bar experimental data.

9.03.3.2.2 Core formation/accretion models

As described in the previous section, siderophile element concentrations are depleted in the Earth's mantle relative to chondritic compositions as a consequence of core formation (**Figures 10** and **11**). However, compared with predicted depletions based on element partitioning studies at 1 bar and moderate temperatures (e.g., 1300–1400 °C), the concentrations of siderophile elements are too high (**Figure 12**; Table 1 in Wood et al., 2006). In the case of MSEs, the discrepancies are around 1–2 orders of magnitude. In the case of the HSEs, the discrepancies are on the order of 5–10 orders of magnitude.

The apparent overabundance of siderophile elements in the mantle has led to a number of core formation hypotheses, most notably the following:

- Heterogeneous two-stage accretion models
- Metal–silicate equilibration at high pressures and temperatures

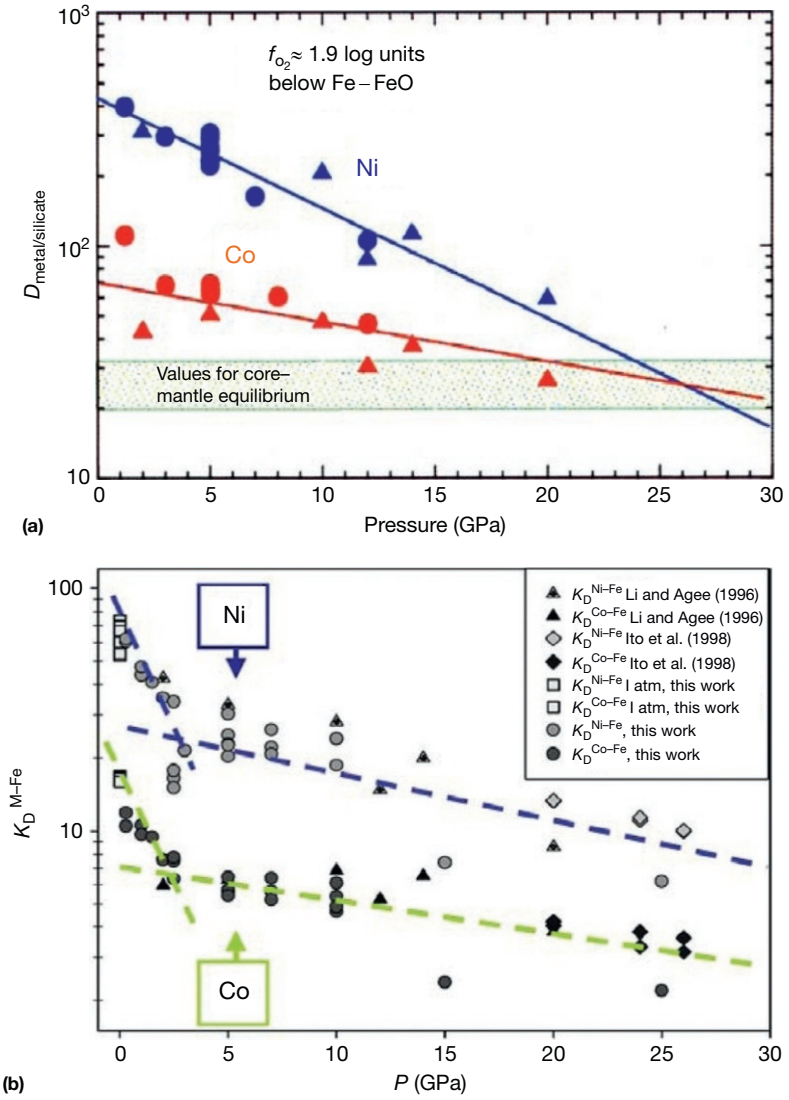


Figure 12 (a) Early experimental results on the effects of pressure on the partitioning of Ni (blue symbols) and Co (red symbols) between liquid Fe alloy and liquid silicate melt, at 2123–2750 K and f_{O_2} = IW-1.9, that led to the hypothesis of metal and silicate equilibrating during core formation at the bottom of a deep magma ocean. Triangles show data from Li and Agee (1996) and filled circles show data from Thibault and Walter (1995) (from Wood et al., 2006). (b) More recent results on the partitioning of Ni and Co (Kegler et al., 2008) show that the pressure dependence undergoes a pronounced change at 3–5 GPa, probably because of pressure-induced structural changes in the silicate liquid (Keppler and Rubie, 1993) and/or the metallic liquid (Sanloup et al., 2011); above 5 GPa, the pressure dependences for both Ni and Co are considerably weaker than the trends shown in (a). Here, $K_D^{M\text{-}Fe}$ is a distribution coefficient in which the partition coefficient has been normalized to Fe partitioning and is thus independent of f_{O_2} (see eqn [10]); in addition, the data have all been normalized to 2273 K by extrapolation. Reproduced from Kegler P, Holzheid A, Rubie DC, Frost DJ, and Palme H (2005) New results of metal/silicate partitioning of Ni and Co at elevated pressures and temperatures. *XXXVI Lunar and Planetary Science Conference*, Abstr. #2030.

- Models of continuous and multistage core formation
- The late-veneer hypothesis
- Addition of core material to the lower mantle

We briefly review each of these hypotheses in turn.

9.03.3.2.2.1 Heterogeneous two-stage accretion models

An early solution to the apparent 'excess siderophile element problem' was proposed by Wänke (1981). His model involved the Earth initially accreting from small bodies that had differentiated under highly reducing conditions at low pressure such that all elements more siderophile than Fe were effectively removed from the silicate mantles into the metallic cores. When these bodies accreted to the Earth, their cores merged with the Earth's protocore without any further chemical equilibration between metal and silicate, thus resulting in the mantle being highly depleted in siderophile elements. During a later second stage, oxidized CI-like material accreted to the Earth and, through the mixing of this material into the Earth's mantle, the current mantle abundances of the siderophile

elements were established. This low-pressure heterogeneous core formation model was further developed by O'Neill (1991) to also explain the concentrations of the HSEs in the mantle.

The evolution from initially highly reducing conditions to later oxidizing conditions during Earth's accretion, proposed by Wänke (1981) and O'Neill (1991), is also an essential feature of more recent core formation models discussed in the succeeding text (Rubie et al., 2011; Wade and Wood, 2005).

However, metal–silicate equilibration only under low-pressure conditions can be excluded on the basis of experimental partitioning data of Mann et al. (2009). This study shows that the chondritic Ga/Mn and ln/Zn ratios of the Earth's mantle require metal–silicate equilibration at an average pressure on the order of 30–60 GPa (see also Figure 1 in Rubie et al. (2011)).

9.03.3.2.2.2 Metal–silicate equilibration at high pressure and temperature

Murthy (1991) proposed that mantle siderophile element abundances could be explained if temperatures during core formation were extremely high. Although there were significant problems with the thermodynamic arguments on which this suggestion was based (Jones et al., 1992; O'Neill, 1992), metal–silicate equilibration at combined high temperatures and pressures may provide the explanation for at least the MSE abundances. This topic has been reviewed in detail by Walter et al. (2000), Righter and Drake (2003), and Wood et al. (2006).

Preliminary high-pressure studies produced results that were inconclusive in determining whether metal–silicate equilibration can explain the siderophile element anomaly (Hillgren et al., 1994; Thibault and Walter, 1995; Walter and Thibault, 1995; Walker et al., 1993). However, an early study of the partitioning of Ni and Co between liquid Fe alloy and silicate melt suggested that the metal–silicate partition coefficients for these elements reach values that are consistent with mantle abundances at a pressure of 25–30 GPa (**Figure 12(a)**; Li and Agee, 1996). This important result led to the idea that metal–silicate equilibration at the base of a magma ocean ~800 km deep can explain the mantle abundances of at least the MSEs (Li and Agee, 1996; Righter et al., 1997). Since the late 1990s, there have been numerous subsequent studies of the partitioning of the MSEs (e.g., Bouhifd and Jephcoat, 2003, 2011; Chabot and Agee, 2003; Chabot et al., 2005; Gessmann and Rubie, 1998, 2000; Kegler et al., 2008; Li and Agee, 2001; Mann et al., 2009; O'Neill et al., 1998; Righter, 2011; Righter and Drake, 1999, 2000, 2001; Siebert et al., 2011, 2012, 2013; Tschauner et al., 1999; Tuff et al., 2011; Wade and Wood, 2005; Wade et al., 2012). A growing consensus emerging from such studies is that the pressures and temperatures required for metal–silicate equilibration were considerably higher than originally suggested. Estimated conditions are quite variable and range up to >4000 K and 60 GPa (**Table 3**; **Figure 13**). One of the reasons for the large scatter of P–T estimates is that, based on the current experimental data set and the associated uncertainties, a wide range of P–T–f_{O_2} conditions can satisfy mantle abundances of the MSEs (**Figure 13**). In addition, the difficulty of identifying a unique set of conditions is hardly surprising considering that core formation occurred over a protracted time period during accretion as the likely consequence of multiple melting events under a range of conditions.

The solubilities of HSEs in silicate liquid have been investigated extensively at 1 bar and moderate temperatures of 1300–1400 °C (e.g., Ertel et al., 1999, 2001; Fortenfant et al., 2003a, 2006), whereas there have been relatively few studies of the solubility or partitioning of HSEs at high pressure. All such studies are beset by a serious technical problem. The quenched samples of silicate liquid inevitably contain numerous metal micronuggets consisting of, or rich in, the element of interest, especially at low f_{O_2}, which make it very difficult to obtain reliable chemical analyses. Although it is normally considered that such nuggets were present in the melt at high temperature (e.g., Holzheid et al., 2000b; Lindstrom and Jones, 1996), it has also been suggested that they form by exsolution from the silicate liquid during quenching (Cottrell and Walker, 2006). However, this latter interpretation can be excluded because Médard et al. (2010) have shown unequivocally, through a high-pressure centrifuge study, that Pt nuggets are present at high temperature and are not the product of quenching. In early studies, the solubilities of Pt and Pd in silicate liquid (Cottrell and Walker, 2006; Ertel et al., 2006; Holzheid et al.,

Table 3 Metal–silicate equilibration conditions during core formation inferred from a selection of studies of siderophile element partitioning

P *(GPa)*	T *(K)*	*References*	*Notes*
28	2400–2700	Li and Agee (1996)	Ni, Co; **T* fixed by peridotite liquidus; $f_{O_2} \approx$ IW-0.5 (expt.)
27	2200	Righter et al. (1997)	Ni, Co, P, Mo, W; f_{O_2} = IW-0.15 (inf.)
37	2300	O'Neill et al. (1998)	Ni, Co, Fe, Cr require 3400 K
>35	>3600	Gessmann and Rubie (2000)	V, Cr, Mn; f_{O_2} = IW-2.3 (inf.)
43–59	2400–4200	Li and Agee (2001)	Ni, Co; f_{O_2} = IW to IW-2 (expt.)
25	3350	Gessmann et al. (2001)	Si; **P* fixed by Ni/Co data; temp exceeds peridotite liquidus
27	2250	Righter and Drake (2003)	P, W, Co, Ni, Mo, Re, Ga, Sn, Cu; f_{O_2} = IW-0.4 (inf.)
40	2800	Walter and Tronnes (2004)	Ni, Co
40	3750	Wade and Wood (2005)	V, Ni, Co, Mn, Si; **T* fixed by peridotite liquidus; evolving f_{O_2}
30–60	>2000	Chabot et al. (2005)	Ni, Co; f_{O_2} = IW-2.2 (inf.); demonstrates solution tradeoffs
22.5	2700	Righter et al. (2010)	11 MSEs; f_{O_2} = IW-2
27–45	3300–3600	Righter (2011)	Compilation of data for 11 MSEs; f_{O_2} = IW-1 to IW-1.5

f_{O_2} is the oxygen fugacity, IW indicates the iron–wüstite buffer, 'inf.' means inferred f_{O_2} for the magma ocean, and 'expt.' indicates the experimental value.

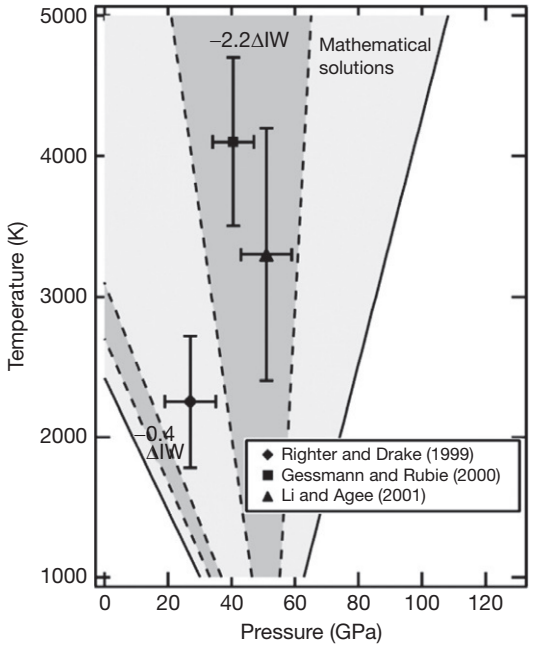


Figure 13 Pressure–temperature conditions of metal–silicate equilibration during core formation that are consistent with the concentrations of Ni and Co in the Earth's mantle, based on experimental studies of the metal–silicate partitioning of these elements. The two dark gray-shaded regions show results calculated for oxygen fugacities of IW-0.4 and IW-2.2, respectively, whereas the light-shaded region shows solutions for all oxygen fugacities. *P–T* estimates from three previous studies are shown by the symbols with error bars. Temperature is poorly constrained because the partition coefficients for Ni and Co depend only weakly on this variable. In order to better constrain the *P–T* conditions of equilibration, it is necessary to consider additional siderophile elements for which partitioning is more strongly temperature-dependent (e.g., Wade and Wood, 2005). Reproduced from Chabot NL, Draper DS, and Agee CB (2005) Conditions of core formation in the Earth: Constraints from nickel and cobalt partitioning. *Geochimica et Cosmochimica Acta* 69: 2141–2151.

2000b), the metal–silicate partitioning of Re (Ohtani and Yurimoto, 1996), the partitioning of Re and Os between magnesiowüstite and liquid Fe (Fortenfant et al., 2003b), and the partitioning of Pd (Righter et al., 2008) were investigated at high pressure. It has been argued that the concentrations of Re, Pt, and Pd in the mantle can be explained by metal–silicate equilibration at high pressure (Cottrell and Walker, 2006; Righter and Drake, 1997; Righter et al., 2008). However, the study of Mann et al. (2012), of the partitioning of Ru, Rh, Pd, Re, Ir, and P between metal and silicate liquids at pressures up to 18 GPa, shows that the concentrations of all HSEs in the mantle and, in particular, their chondritic ratios cannot be explained by high-pressure equilibration, as concluded also by the majority of the studies cited earlier (see also Brenan and McDonough, 2009). In addition, based on studies of Martian meteorites, HSE abundances are similar in the mantles of both Earth and Mars (Walker, 2009; Warren et al., 1999). This cannot be explained by metal–silicate equilibration because, in planets of dissimilar sizes, *P–T* conditions of core formation must be quite different. Therefore, metal–silicate HSE partitioning results are generally interpreted to support the late-veneer hypothesis (see next section) for both Earth and Mars (e.g., Day et al., 2012; Righter, 2005).

There are serious problems with the simple concept of metal–silicate equilibration at the base of a magma ocean (Rubie et al., 2003) and, in particular, with the so-called 'single-stage' core formation models as proposed, for example, by Corgne et al. (2009) and Righter (2011). The complete process of core formation in the Earth cannot have been accomplished physically through a single event involving a magma ocean of limited depth (e.g., 800–1500 km). Several large impacts are likely to have occurred during accretion, each generating a magma ocean with different characteristics, while there may also have been a steady background flux of smaller impactors.

Each impact would have caused an episode of core formation, with conditions varying from one impact to the next. The claim of Righter (2011) that "The single PT point of this study is likely the last record of major equilibration in this series of large magnitude impacts and subsequent melting leading to the Earth's final size. The energy associated with a large impact and subsequent heating due to metal–silicate segregation, will cause extensive reequilibration" requires that the entire core and mantle of the Earth reequilibrated chemically at a depth of 700–1200 km after accretion was almost complete. Because of the large density difference between core and mantle and the need for the core to fully emulsify, such reequilibration is physically impossible, especially at a relatively shallow depth. Thus, the conditions listed in **Table 3** must be considered, at best, to be average values of a broad range of conditions.

9.03.3.2.2.3 Models of continuous and multistage core formation

In order to study core formation more realistically, two approaches have been adopted, involving (a) continuous core formation (e.g., Wade and Wood, 2005) and (b) multistage core formation (Rubie et al., 2011, 2015). Both models have an important feature in common with the heterogeneous two-stage accretion models described earlier – a transition from initially very reducing to more oxidizing conditions during accretion. However, they differ from the earlier models by requiring metal–silicate equilibration at high pressures.

Wade and Wood (2005) presented a model in which core formation occurs continuously during accretion (see also, e.g., Tuff et al., 2011; Wade et al., 2012; Wood et al., 2006). This model utilizes the metal–silicate partitioning of a range of elements, including Fe, Ni, Co, V, W, Nb, and Cr, and is based on the assumption of metal–silicate equilibration at or near the base of a magma ocean that deepens as the Earth grows in size. The temperature at the base of the magma ocean is constrained to lie on the peridotite liquidus. Wade and Wood (2005) discovered that, in order to satisfy the observed mantle abundances of all elements considered (and especially V), it is necessary that the oxygen fugacity is initially very low but increases during accretion to satisfy the current FeO content of the mantle (**Figure 14**). Typically in such models, equilibration pressures increase up to a maximum of 40–50 GPa during accretion (e.g., **Figure 14(b)**).

The exact oxygen fugacity path followed during accretion cannot be defined uniquely in these models (e.g., see **Figure 14(a)**). The original explanation for the increase in f_{O_2}

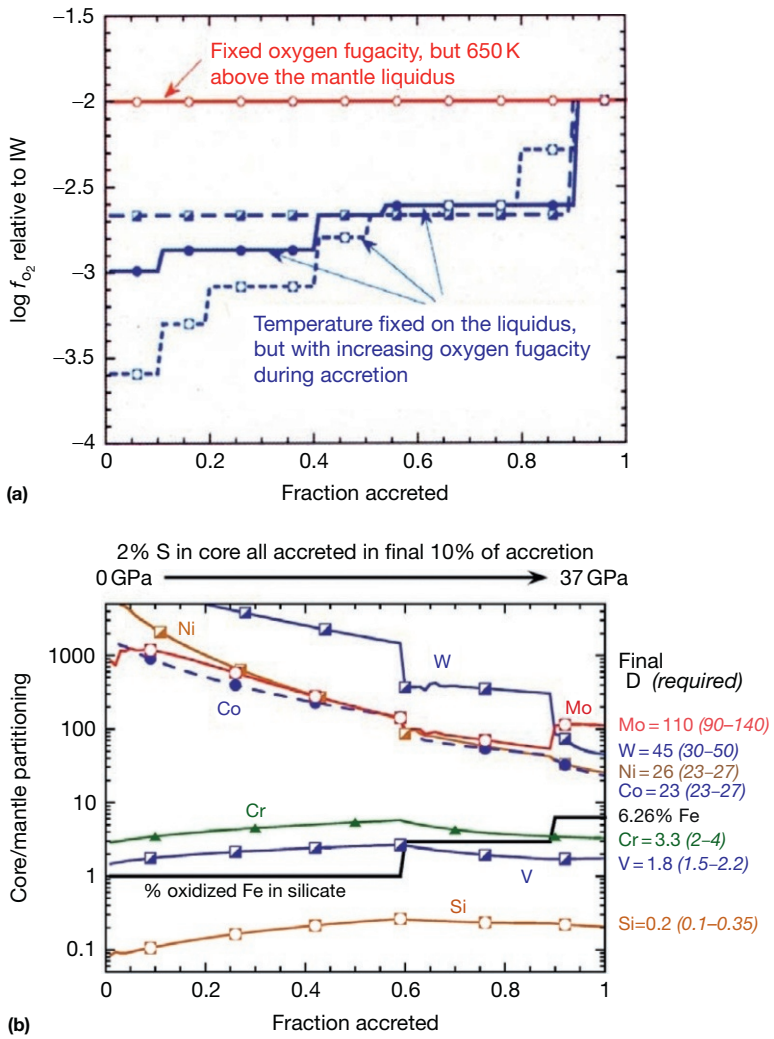


Figure 14 Models involving continuous core formation during accretion of the Earth based on metal–silicate equilibration at the base of a deepening magma ocean. (a) Oxygen fugacity is plotted as a function of fraction accreted. The temperature at the base of the magma ocean is constrained to lie on the peridotite liquidus for the variable oxygen fugacity models. In order to satisfy concentrations of a range of siderophile elements in the mantle, the oxygen fugacity has to be low initially and then increase during accretion for reasons described in the text. Several possible oxygen fugacity evolution paths are shown (Reproduced from Wood BJ, Walter MJ, and Wade J (2006) Accretion of the Earth and segregation of its core. *Nature* 441: 825–833. http://dx.doi.org/10.1038/nature04763.). (b) The evolution of core–mantle partition coefficients for Si, Fe, V, Cr, Ni, Co, W, and Mo as a function of the fraction of the Earth accreted. The addition of S to the core during the final stages of accretion enables the final values for W and Mo to be achieved (Reproduced from Wade J, Wood BJ, and Tuff J (2012) Metal–silicate partitioning of Mo and W at high pressures and temperatures: Evidence for late accretion of sulphur to the Earth. *Geochimica et Cosmochimica Acta* 85: 58–74.).

is based on the so-called 'oxygen pump' mechanism. This involves the crystallization of silicate perovskite from the magma ocean, which can only occur once the Earth has reached a critical size such that the pressure at the base of the magma ocean reaches ~24 GPa. Because of its crystal chemistry, the crystallization of silicate perovskite causes ferrous iron to dissociate to ferric iron + metallic iron by the reaction

$$\underset{\text{silicate liquid}}{Fe^{2+}} \rightarrow \underset{\text{perovskite}}{Fe^{3+}} + \underset{\text{metal}}{Fe}$$

(Frost et al., 2004). If some or all of the metal produced by this reaction is transferred from the mantle to the core, the Fe^{3+} content and the f_{O_2} of the mantle both increase. Wade and Wood (2005), however, incorrectly argued that this 'oxygen pump' mechanism could have increased the f_{O_2} of the mantle during core–mantle equilibration from the reduced levels (<ΔIW-3) required in the early stages of their model, where Si and V would have entered the core, to the more oxidized levels (~ΔIW-2) that allowed elements such as Fe, Ni, and Co to partially remain in the mantle. This is not possible because raising the f_{O_2} from this level requires the FeO content of the silicate mantle to increase, whereas the 'oxygen pump' mechanism can only raise the Fe^{3+} content (i.e., the Fe^{3+}/Fe^{2+} ratio) of the mantle. Although the oxygen content of the mantle is increased through disproportionation, the total Fe content of the mantle is actually lowered as Fe must be lost to the core. Although disproportionation raises the O/Fe ratio of the silicate mantle, further reaction of the newly liberated oxygen with

more accreting Fe metal only leaves the FeO content of the mantle at the same level as it was before disproportionation occurred. While the 'oxygen pump' mechanism is effective at raising the Fe^{3+} content of the mantle and the f_{O_2} of the upper mantle to levels above those of iron saturation, it cannot influence the f_{O_2}, while the mantle remains at iron saturation, that is, while core formation was still occurring (see also Frost et al., 2008).

Other possible causes of an increase in f_{O_2} are (a) the partitioning of Si into core-forming metal, which increases the FeO content of the mantle (see succeeding text); (b) the addition of bodies to the Earth that are increasingly oxidized as accretion progresses (see also succeeding text); and (c) the dissociation of accreted water due to hydrogen loss.

A model of multistage core formation has been formulated by Rubie et al. (2011). This involves the Earth accreting through a series of violent collisions with smaller differentiated bodies, with each collision resulting in a magma ocean and an episode of metal–silicate segregation. The model is based on a simple and idealized accretion scenario whereby each impacting body has a mass around 10% of the Earth's mass at the time of impact. Thus, as the Earth grows in size, the impacting bodies also become larger. The metallic cores of the impactors fully or partially equilibrate in the magma ocean before merging with the Earth's protocore. It is further assumed that the metal–silicate equilibration pressure is a constant fraction of the CMB pressure at the time of impact and that the corresponding equilibration temperature lies on the peridotite liquidus. In contrast to previous core formation models, no assumptions are made about oxygen fugacity. Instead, the bulk composition of accreting material is defined in terms of the nonvolatile elements Si, Al, Mg, Fe, Al, Ni, Co, Cr, V, Nb, W, and Ta. These elements are considered to have solar system (CI chondritic) relative abundances but with enhanced concentrations of the refractory elements Ca, Al, V, Nb, W, and Ta (Palme and O'Neill, 2003). Different bulk oxygen contents result in different oxidation states. For example, an oxygen-poor (reduced) composition would result in a high proportion of total Fe being present as metal and a significant amount of Si being dissolved in the metal. A more oxidized composition would have a significant proportion of total Fe as FeO and zero Si in the metal. In order to determine the compositions of coexisting silicate and metal liquids after equilibration at high pressure and temperature, the compositions are expressed in terms of the molar concentrations of major elements, as follows:

$$\underset{\text{silicate liquid}}{\left[(\mathrm{FeO})_x(\mathrm{NiO})_y(\mathrm{SiO_2})_z(\mathrm{Mg}_u\mathrm{Al}_m\mathrm{Ca}_n)\mathrm{O}\right]} + \underset{\text{metal liquid}}{[\mathrm{Fe}_a\mathrm{Ni}_b\mathrm{O}_c\mathrm{Si}_d]}$$

Here, the indexes u, m, and n, which describe the concentrations of lithophile elements, are determined from the bulk composition. The other seven indexes, x, y, z, a, b, c, and d, are determined using four mass balance equations (for Fe, Ni, Si, and O), based on the bulk composition, and three models that describe the metal–silicate partitioning of Ni, Si, and O. The concentrations of the trace elements Co, Cr, V, Nb, W, and Ta have a negligible effect on the mass balance and are determined from partitioning alone. As mentioned earlier, this approach avoids making any assumptions about oxygen fugacity, although this parameter can be determined from the partitioning of Fe (e.g., **Figure 2** in Rubie et al., 2011).

Using this approach to model core formation, the evolving compositions of the mantle and core were determined as a function of mass accreted (**Figure 15**). The main constraint is the composition of the Earth's primitive mantle (e.g., Palme and O'Neill, 2003), and this composition was achieved in the model through a least squares optimization of several fitting parameters. The latter include the oxygen content(s) of accreting material, the metal–silicate equilibration pressure, and the fraction of metal that equilibrates in the magma ocean. Poor results were obtained for homogeneous accretion – that is, when the oxygen content of the accreting material is constant. Best results were obtained with heterogeneous accretion with the initial 60–70% of the Earth accreting from highly reduced material and the final 30–40% accreting from more oxidized material. This scenario is consistent with accretion models because material originating relatively close to the Sun (e.g., ~1 AU) is expected to be volatile-poor and reduced, while material originating far from the Sun would be richer in volatiles (including water) and therefore more oxidized. The evolution from reduced to oxidized compositions results in an oxygen fugacity evolution (**Figure 15(c)**) that is qualitatively similar to those of the models of Wänke (1981), O'Neill (1991), and Wade and Wood (2005) (see preceding text). It is also consistent with conclusions regarding heterogeneous accretion based on Pd–Ag isotopes (Schönbächler et al., 2010). The initially reduced composition is required in order to partition sufficient V, Cr, and Si into the core, whereas the later addition of oxidized material enables the FeO content of the mantle to be achieved. In addition to requiring heterogeneous accretion, some degree of disequilibrium (incomplete equilibration of the impactors' metallic cores) was found to be necessary in this model.

In addition to the effects of the late accretion of oxidized material, a progressive increase in oxygen fugacity results from the following reaction (originally proposed by Javoy, 1995):

$$\underset{\text{metal}}{2\mathrm{Fe}} + \underset{\text{silicate}}{\mathrm{SiO_2}} = \underset{\text{metal}}{\mathrm{Si}} + \underset{\text{silicate}}{2\mathrm{FeO}}$$

As temperature increases, Si partitions increasingly into the metal (e.g., Mann et al., 2009; see also **Section 9.03.3.3**) with the result that the FeO content of the silicate and f_{O_2} both increase.

9.03.3.2.2.4 The 'late-veneer' hypothesis

According to this hypothesis, the main stage of core formation resulted in the almost complete extraction of HSEs from the mantle into the metallic core, under reducing oxygen fugacity conditions (e.g., Kimura et al., 1974; Morgan, 1986; O'Neill, 1991; O'Neill and Palme, 1998; Righter, 2005; Walker, 2009). At a very late stage of accretion, a thin veneer of chondritic material was added to the Earth under relatively oxidizing conditions, such that the HSEs were retained in the mantle in approximately chondritic ratios. The mass of material added at this late stage is considered to be <1% of the entire mantle. This is currently the most widely accepted 'heterogeneous' core formation model (e.g., Mann et al., 2012; Walker, 2009). An

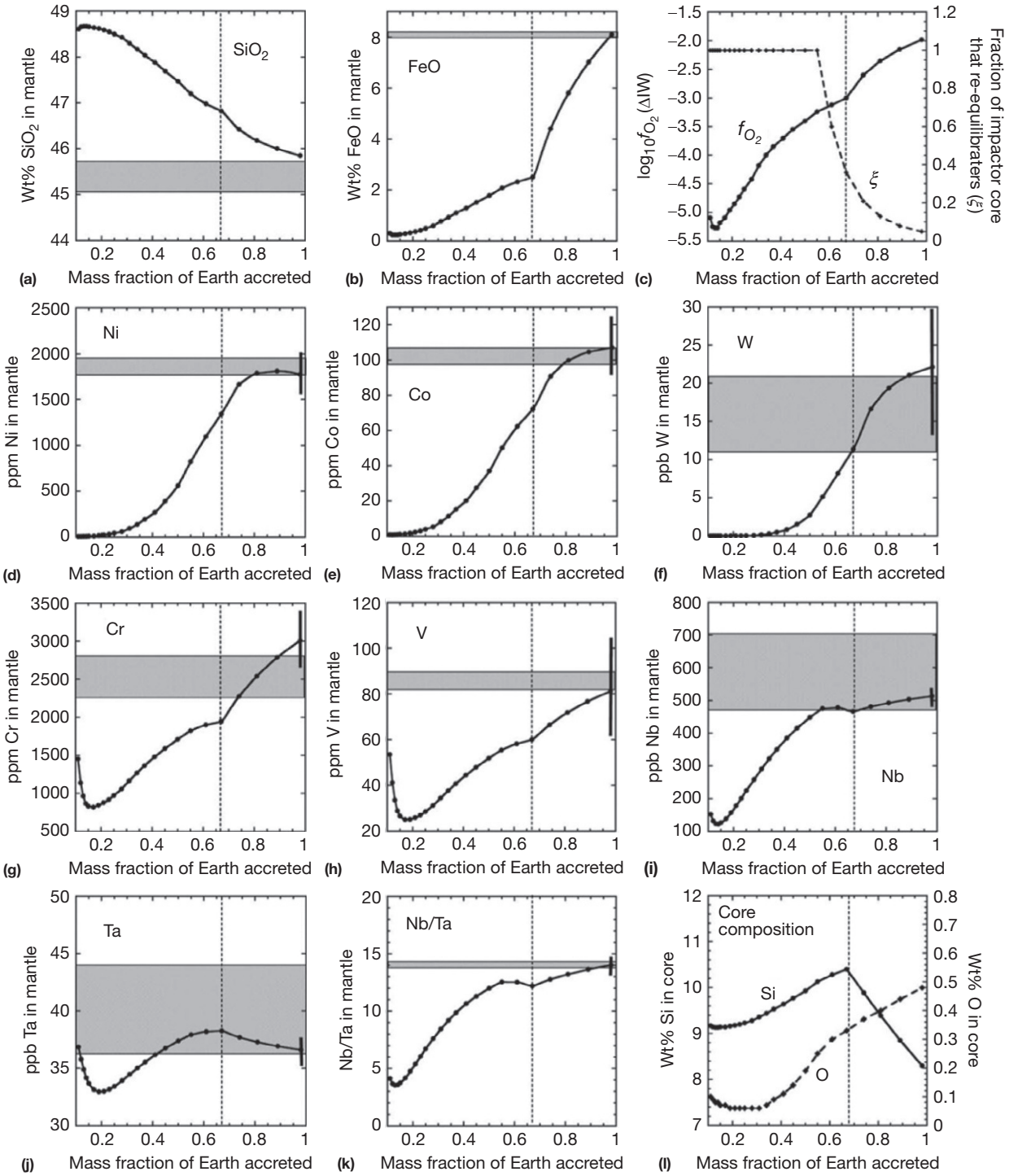


Figure 15 The evolution of mantle (a and b, d–k) and core (l) compositions as a function of mass accreted in the multistage core formation model of Rubie et al. (2011). The evolution of f_{O_2} and the extent of metal–silicate disequilibrium are shown in (c). Reproduced from Rubie DC, Frost DJ, Mann U, et al. (2011) Heterogeneous accretion, composition and core–mantle differentiation of the Earth. *Earth and Planetary Science Letters* 301: 31–42, http://dx.doi.org/10.1016/j.epsl.2010.11.030.

important constraint on this process is that, while the Earth's abundance of HSEs requires a late addition of undifferentiated material, the lack of HSEs intrinsic to the Moon indicates that it did not receive the same addition. Different arguments have been put forth to explain this puzzle, one of the most recent being that the late veneer was added in a collision with a single, large body that struck the Earth but, by the statistics of small numbers, happened to miss the Moon. The dynamics of this late accretion process have been discussed by Bottke et al. (2010).

9.03.3.2.2.5 Addition of outer core material to the lower mantle

It has been suggested that HSE abundances in the Earth's mantle have resulted from core–mantle interaction (Brandon and Walker, 2005), for example, by the addition of a small amount of core metal to the mantle (Snow and Schmidt, 1998). This could potentially occur by capillary action (considered highly unlikely by Poirier et al., 1998) or by dilatancy caused by volume strain (Kanda and Stevenson, 2006; Rushmer et al., 2005). Alternatively, siderophile elements could be added to the base of the mantle by crystallization of oxides or silicates due to the growth of the inner core (Walker, 2000, 2005), and chemical exchange could be facilitated by the presence of partial melt in the lower few kilometers of the mantle or a basal magma ocean (Labrosse et al., 2007). In this case, the metal–silicate partition coefficients at CMB conditions would have to be orders of magnitude lower than at low P–T conditions and would need to result in the addition of the HSEs in approximately chondritic proportions.

9.03.3.2.3 Metal–silicate fractionation models

The simplest model of metal–silicate fractionation during core formation involves the ponding of liquid iron at the base of a convecting magma ocean with chemical equilibration occurring at the metal–silicate interface (**Section 9.03.3.2.2.1**, **Figure 8(a)**). An appealing feature of this model is that siderophile element abundances in the mantle can be interpreted directly in terms of magma ocean depth (e.g., Li and Agee, 1996; Righter, 2011; Righter et al., 1997). The timescale required for chemical equilibration across the metal–silicate interface has been investigated by Rubie et al. (2003) assuming that mass transport occurs by chemical diffusion through boundary layers that exist above and below the interface. Assuming that a magma ocean crystallizes from the bottom up, the initial crystallization of silicate minerals at the base of the magma ocean will terminate equilibration between the ponded iron and the overlying magma ocean. Thus, Rubie et al. (2003) also calculated the timescale required for the initial crystallization of the base of a magma ocean. Results are dependent upon the depth of the magma ocean and suggest that equilibration times are almost three orders of magnitude greater than cooling times (**Figure 16**). Such a result is also predicted by considering that rates of conductive heat transfer are much faster than rates of chemical diffusion (Rubie et al., 2003). Therefore, these results appear to rule out simple chemical equilibration at the base of the magma ocean.

The equilibration model of Rubie et al. (2003) is based on the assumption that a dense atmosphere was not present during magma ocean crystallization. The effect of such an insulating atmosphere would be to reduce the rate of heat loss and therefore prolong the lifetime of the magma ocean. However, the rate of convection would also be reduced, which would slow the rate of chemical exchange across the metal–silicate interface. Thus, the presence of an insulating atmosphere is unlikely to affect the conclusions of Rubie et al. (2003). However, as described earlier, there has been a suggestion that a magma ocean adiabat is much steeper than formerly believed with the consequence that terrestrial magma oceans might crystallize from the top down rather than from the bottom up (Mosenfelder et al., 2007, 2009; c.f. Miller et al., 1991). In this case, magma ocean crystallization would be much slower, and the possibility of simple metal–silicate equilibration at its base might need to be reconsidered.

It is currently assumed in core formation models that metal–silicate separation takes place in a magma ocean of global extent and of more or less constant depth (**Figure 8**). As discussed earlier in **Section 9.03.2.3.2**, this assumption requires reconsideration. **Figure 17** shows an alternative model in which the magma ocean that is generated by a large impact is initially a hemispheric body of limited lateral extent. The attainment of isostatic equilibrium eventually results in the formation of a global magma ocean, but metal segregation could already have taken place before this developed. Pressure at the base of the initial hemispheric magma ocean is clearly much higher than at the base of the final global magma ocean,

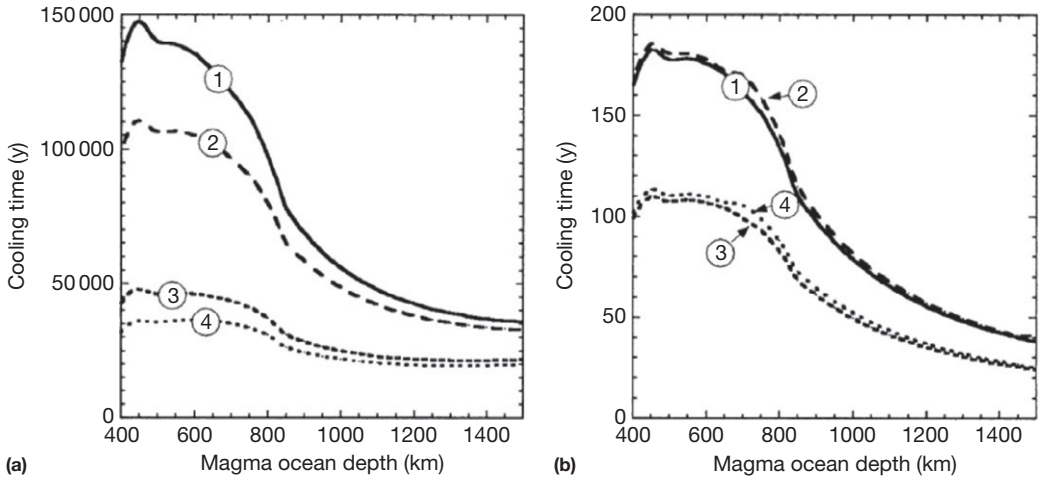


Figure 16 Results of a model in which a layer of segregated (ponded) liquid iron equilibrates with an overlying convecting magma ocean. (a) Time to reach 99% equilibration between metal and silicate as a function of magma ocean depth. (b) Time required for initial crystallization at the base of the magma ocean (thus effectively terminating the equilibration process). The four curves (1–4) are results for a wide range of plausible model parameters. Reproduced from Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

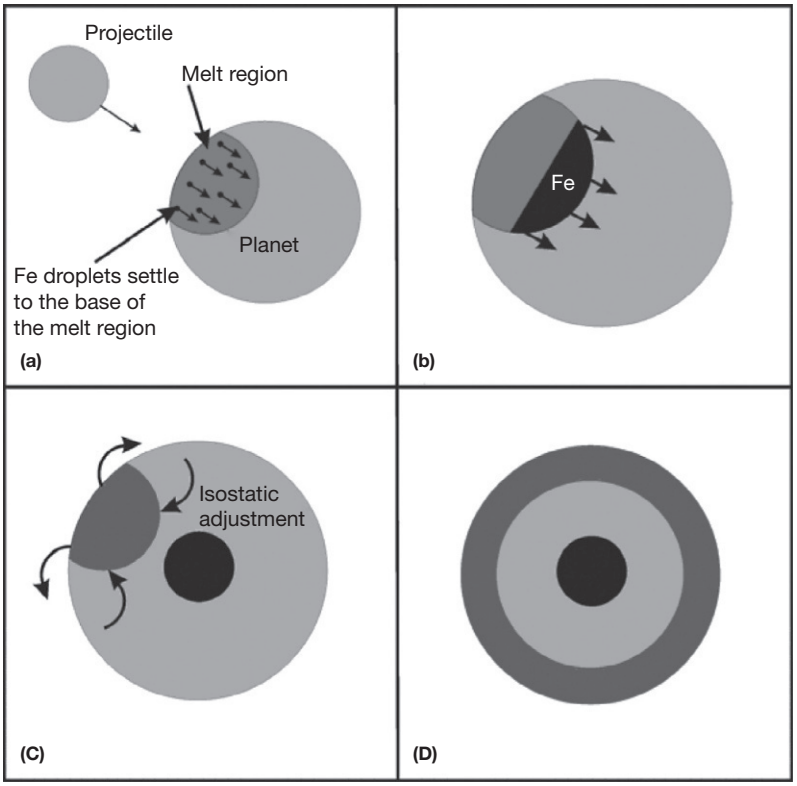


Figure 17 Possible evolution of a deep magma ocean that forms as a consequence of a giant impact. (a) Initially, a hemispheric magma ocean forms that is of limited lateral extent. There is the possibility that iron, emulsified in the form of small dispersed droplets, settles and segregates rapidly to form a protocore at the bottom of this magma ocean (b and c). Subsequent isostatic adjustment causes the magma ocean to evolve into a layer of global extent (d). The relative timing of the processes of metal segregation and isostatic adjustment determines the pressure (and probably also temperature) conditions of metal–silicate equilibration. Reproduced from Tonks WB and Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100: 326–346.

which means that the partitioning of siderophile elements could depend critically on the timing of metal segregation.

As discussed earlier, liquid metal is likely to be present in a magma ocean in the form of small droplets ~1 cm in diameter. Such droplets remain in chemical equilibration with the magma as they settle out and *P*–*T* conditions change because diffusion distances are short (Karato and Murthy, 1997; Rubie et al., 2003). It is necessary to understand the chemical consequences of the settling out of such droplets in order to interpret siderophile element geochemistry in terms of magma ocean depths. Rubie et al. (2003) considered two end-member models for the partitioning of Ni. The models are based on the parameterized Ni partitioning formulation of Righter and Drake (1999), which requires a magma ocean depth of ~800 km for metal–silicate equilibration at its base to produce the estimated core–mantle partition coefficient of ~28. In Model 1 of Rubie et al. (2003), the magma ocean is static (i.e., no convection or mixing) and droplets that are initially uniformly dispersed, settle out, and reequilibrate progressively as they fall. This polybaric equilibration model requires a magma ocean ~1400 km deep (**Figure 18**) because much of the silicate equilibrates at relatively low pressures $D_{\mathrm{M}}^{\mathrm{met-sil}}$ where values are high. Model 2 of Rubie et al. (2003) is based on an assumption of vigorous convection keeping the magma ocean fully mixed and chemically homogeneous. The iron droplets equilibrate finally with silicate liquid at the base of the magma ocean just before segregating into a ponded layer. Because the mass fraction of metal that is available to equilibrate with silicate progressively decreases with time, the effectiveness of metal to remove siderophile elements from the magma ocean also decreases. This model requires a magma ocean ~550 km deep to produce the desired core–mantle partition coefficient of 28 (**Figure 18**).

The two end-member Models 1 and 2 are clearly both physically unrealistic, and a more realistic model must lie somewhere between the two extremes shown in **Figure 18**. In order to investigate the chemical consequences of metal–silicate segregation in a deep magma ocean more rigorously, Höink et al. (2006) have combined two- and three-dimensional numerical convection models with a tracer-based sedimentation method. They found that metal droplets stabilize the magma ocean against convection and that convection only develops in the upper layer of a magma ocean after it has become depleted in iron droplets. The timescale for droplet separation was found to be identical to the Stokes' settling time. Because of the lack of convection in the droplet-dominated layer, the chemical consequences are similar to those of Model 1 of Rubie et al. (2003), although the results of the two- and three-dimensional models of Höink et al. (2006) appear to be very different from each other (see their Figure 20). It must also be noted that, for computational reasons, the Rayleigh numbers in the models

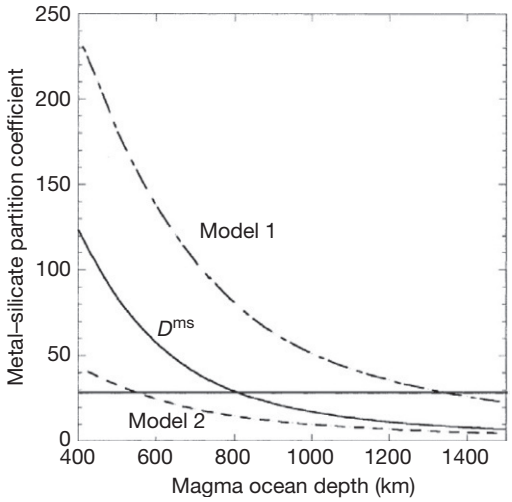


Figure 18 Results of different metal–silicate fractionation models for the siderophile element Ni as a function of magma ocean depth. The core–mantle partition coefficient for this element is estimated to be ~28 (as indicated by the horizontal line). The curve labeled D^{ms} indicates results for simple equilibration between a ponded liquid metal layer and the overlying magma ocean and requires a magma ocean ~800 km deep to obtain the core–mantle value. However, this model is unlikely to be realistic (**Figure 16**). Models 1 and 2 represent extreme end-member cases (see text for details) and require magma ocean depths of ~1400 and ~550 km, respectively. Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003). Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

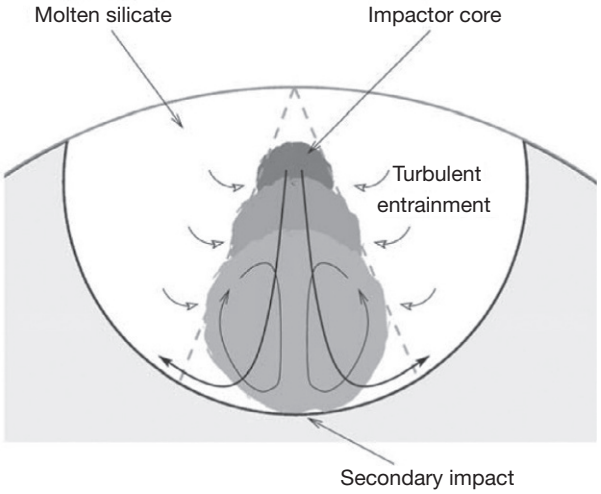


Figure 19 Model of a metallic core of an impactor sinking in a magma ocean. The turbulent entrainment of silicate liquid leads to the descent of a metal–silicate plume. With increasing depth, the plume broadens and an increasing volume of silicate is entrained. Chemical equilibration between metal and the entrained silicate liquid is expected to occur near the base of the magma ocean. Reproduced from Deguen R, Olson P, and Cardin P (2011) Experiments on turbulent metal–silicate mixing in a magma ocean. *Earth and Planetary Science Letters* 310: 303–313.

of Höink et al. (2006) have values of $\leq 10^7$, whereas the Rayleigh number for a deep magma ocean is on the order of 10^{27}–10^{32}, that is, 20–25 orders of magnitude higher (see **Section 9.03.2.3.2**). It is also difficult to understand the lack of composition-driven convection in the Höink et al. model. Only slight variations in the iron/silicate mixture ratio in the lower part of the ocean should create density variations much larger than those induced by thermal variations and thus drive a much more vigorous convection that, while the iron is settling, may overwhelm thermal convection. The Höink et al. models also seem to neglect the very substantial heating that may occur in the lower part of the magma ocean due to the conversion of the gravitational energy of the falling iron droplets into thermal energy, as discussed in **Section 9.03.2.2.3**. Preliminary numerical models of the iron separation process that do include these effects give substantially different results than the Höink et al. models (Melosh and Rubie, 2007). This is a rapidly developing area of research and controversy at the current time.

An alternative and perhaps more realistic physical model for the descent and equilibration of an impactor's core through a magma ocean has been presented by Deguen et al. (2011). This involves the silicate liquid immediately adjacent to the descending mass of metal undergoing turbulent entrainment such that a mixture of metal and silicate descends as an expanding plume or density current (**Figure 19**). The radius of the entrained silicate plume increases linearly with depth so that the volume ratio of metal to silicate in the descending plume decreases with depth and can be determined from

$$\phi = \left(\frac{r_0}{r}\right)^3 = \left(1 + \frac{\alpha z}{r_0}\right)^{-3} \quad [12]$$

where ϕ is the volume fraction of metal in the metal–silicate mixture, r_0 is the original radius of the impacting core, r is the radius of the descending plume, z is depth, and α is a constant with a value ~0.25. An important consequence of this model is that the metal only equilibrates chemically with a small fraction of the total volume of silicate melt in the magma ocean – which, in turn, makes the equilibration pressures required to reproduce mantle siderophile element concentrations higher.

9.03.3.2.4 Concluding remarks

The consistency of MSE abundances with metal–silicate equilibration at high P–T conditions implies that element abundances produced by earlier equilibration in smaller bodies at lower P–T conditions must have been strongly overprinted. This conclusion supports the idea of impactor cores undergoing reequilibration in the magma ocean, presumably as a result of partial or complete emulsification (Rubie et al., 2003). Later events that did not involve reequilibration (e.g., the descent of large iron diapirs through the lower mantle) would leave little or no signature in the observed siderophile element abundances.

9.03.3.3 Light Elements in the Core

The density of the Earth's core is too low, by 5–10%, for it to consist only of Fe and Ni (e.g., Anderson and Isaak, 2002; Birch, 1952). It has therefore been postulated that the core must contain up to ~10 wt% of one or more light elements, with the most likely candidates being S, O, Si, C, P, and H (Poirier, 1994). Knowledge of the identity of the light element(s) is important for constraining the bulk composition

of the Earth and for understanding processes occurring at the CMB and how such processes are affected by crystallization of the inner core (e.g., Buffett et al., 2000; Buffett and Seagle, 2010; Helffrich and Kaneshima, 2004). The main constraints on the identity of the light elements present in the core are based on cosmochemical arguments (McDonough, 2003), experimental data (Badro et al., 2007; Hillgren et al., 2000; Li and Fei, 2003), computational simulations (e.g., Alfe et al., 2002), silicon isotope fractionation (e.g., Fitoussi et al., 2009; Georg et al., 2007; Ziegler et al., 2010), and core formation modeling (Rubie et al., 2011). The topic has been reviewed in part by Hillgren et al. (2000), McDonough (2003), and Li and Fei (2003), and here, we provide only a brief summary of the main arguments and experimental results.

The sulfur content of the core, based on the volatility of this element, is likely to be no more than 1.5–2 wt% (Dreibus and Palme, 1996; McDonough, 2003; McDonough and Sun, 1995). Similarly, cosmochemical constraints suggest that only very small amounts (e.g., $\leq$0.2 wt%) of C and P are present in the core (McDonough, 2003); therefore, these elements are unlikely to contribute significantly to the density deficit.

One of the main controversies concerning the identity of the principal light element(s) in the core involves silicon and oxygen. A high oxygen content is favored by high oxygen fugacity, whereas a high Si content is favored by low oxygen fugacity (Hillgren et al., 2000; Kilburn and Wood, 1997; Li and Fei, 2003; Malavergne et al., 2004). It has therefore been proposed that these two elements are almost mutually exclusive (Figure 10 in O'Neill et al., 1998). Some models of core composition have been based on this exclusivity – that is, it is assumed that the core contains either Si or O but not both of these elements (McDonough, 2003, Table 7). However, the effects of high pressure and temperature are also critical.

Based on experimental results of Gessmann et al. (2001), obtained up to 23 GPa and 2473 K, the solubility of Si in liquid iron increases with both P and T. An extrapolation of their experimental data shows that ~7 wt% Si can be dissolved in liquid Fe at 25–30 GPa, 3100–3300 K, and an f_{O_2} (IW-2), conditions that are often considered to be applicable for core formation (**Figure 20**). This result is in accordance with the geochemical model of the core of Allegre et al. (1995) in which ~7 wt% Si in the core was proposed based on the Mg/Si ratio of the mantle being high compared with CI chondrites. It is also consistent with the results of Si isotope studies (Fitoussi et al., 2009; Georg et al., 2007; Ziegler et al., 2010). Experimental results comparable to those of Gessmann et al. (2001), suggesting a strong effect of temperature and a weak effect of pressure on the metal–silicate partitioning of Si, have been published by Wade and Wood (2005) and Mann et al. (2009).

Oxygen was first proposed as the principal light element in the core more than 30 years ago (e.g., Ohtani and Ringwood, 1984; Ohtani et al., 1984; Ringwood, 1977). Whereas it is clear that the solubility of this element in liquid Fe increases strongly with temperature, the effect of pressure has been controversial. According to studies of phase relations in the Fe–FeO system, solubility increases with pressure (Kato and Ringwood, 1989; Ringwood, 1977). However, investigations of the partitioning of FeO between magnesiowüstite and liquid Fe have indicated that solubility decreases with increasing pressure and, based on a very large extrapolation, would be essentially zero at

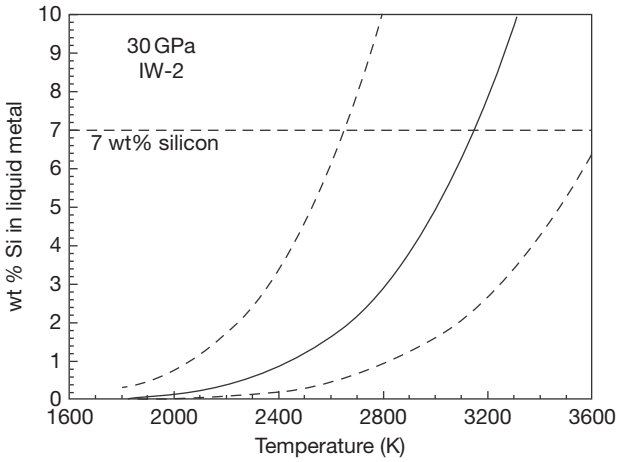


Figure 20 Solubility of Si in liquid Fe alloy at 30 GPa and an oxygen fugacity of IW-2 based on an extrapolation of experimental data. The dashed lines show the extent of the uncertainty envelope based on error propagation. Reproduced from Gessmann CK, Wood BJ, Rubie DC, and Kilburn MR (2001) Solubility of silicon in liquid metal at high pressure: Implications for the composition of the Earth's core, *Earth and Planetary Science Letters* 184: 367–376.

conditions of the CMB (O'Neill et al., 1998; Rubie et al., 2004). Asahara et al. (2007) and Frost et al. (2010) have resolved the controversy by showing that the partitioning of FeO into liquid Fe decreases weakly with pressure up to ~15 GPa and then increases again at high pressures. The reason for the change in pressure dependence is that the Fe–O component dissolved in liquid Fe is more compressible than FeO dissolved in an oxide or silicate phase. Thus, the sign of the volume change (ΔV) of the exchange reaction of oxygen between silicates/oxides and liquid Fe reverses at 10–15 GPa (see also Ohtani et al., 1984; Walker, 2005). As discussed in the succeeding text, the partitioning of oxygen into liquid iron appears to be high enough at core formation conditions for this element to be a significant light element in the Earth's core (Asahara et al., 2007; Rubie et al., 2004; Siebert et al., 2013).

High-pressure studies using the LH-DAC suggest that concentrations of both oxygen and silicon could be significant in the core and could, in combination, account for the density deficit. Takafuji et al. (2005) found 3 wt% Si and 5 wt% O in liquid Fe in equilibrium with (Mg,Fe)SiO3 perovskite at 97 GPa and 3150 K. In similar experiments on liquid Fe coexisting with the postperovskite phase, Sakai et al. (2006) found up to 6.3 wt% O and 4 wt% Si in liquid Fe at 139 GPa and 3000 K. These results should be regarded as preliminary because of the huge temperature gradients and large temperature uncertainties that are characteristic of LH-DAC experiments; in addition, the oxygen fugacity was probably not buffered. In the case of oxygen, the LH-DAC results are mostly consistent with the lower-pressure data of Asahara et al. (2007). However, the silicon results indicate considerably higher solubilities than those predicted by Gessmann et al. (2001). This may be due to Si–O interactions that cause the partitioning of Si into metal to be enhanced by the presence of oxygen in the metal (Tsuno et al., 2013). The high mutual solubilities of O and Si, compared with low P–T results

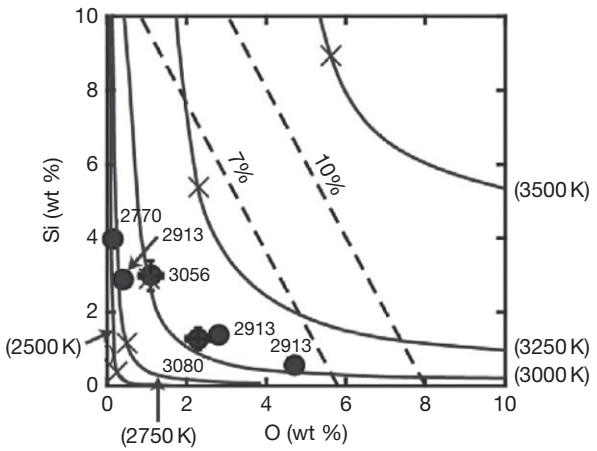


Figure 21 Concentrations of O and Si in liquid iron in equilibrium with silicate liquid at 25 GPa based on experimental data (filled circles) obtained up to 3080 K; the solid lines show a fitted thermodynamic model calculated at temperatures up to 3500 K (Tsuno et al., 2013). (For comparison, see a similar 1 bar plot of O'Neill et al., 1998, **Figure 10**.) Because high O concentrations are favored by high f_{O_2} and high Si concentrations require a low f_{O_2}, the concentrations of these elements in liquid Fe are almost mutually exclusive at low temperatures. At temperatures >3000 K, however, both elements can be present in significant concentrations. The dashed lines indicate concentrations required to account for the 7–10% density deficit of the core, and the crosses indicate concentrations that would result from equilibration with silicate liquid containing 8 wt% FeO. Reproduced from Tsuno K, Frost DJ, and Rubie DC (2013) Simultaneous partitioning of silicon and oxygen into the Earth's core during early Earth differentiation. *Geophysical Research Letters* 40: 66–71.

(O'Neill et al., 1998), are largely the consequence of high temperatures as shown in **Figure 21**.

An additional important observation is that the outer core has a larger density deficit and therefore appears to contain a higher concentration of light element(s) than the inner core (Alfe et al., 2002; Jephcoat and Olson, 1987). This implies that the light element(s) partitions strongly into liquid iron during freezing, which is potentially diagnostic behavior. For instance, Alfe et al. (2002), using molecular dynamics simulations, found that oxygen, due to its small atomic radius, tends to be expelled during the freezing of liquid iron. Conversely, S and Si have atomic radii similar to that of iron at core pressures and thus substitute freely for iron in the solid inner core. Based on the observed density difference between the inner and the outer cores, Alfe et al. (2002) predicted ~8 wt% oxygen in the outer core. These results support the case for O being the main light element and are in agreement with the predictions of Asahara et al. (2007) and Siebert et al. (2013).

The Earth's core is considered not to be in chemical equilibrium with the mantle (Stevenson, 1981), and light element solubilities at core conditions are therefore not necessarily indicative of the actual concentrations of these elements in the core. Instead, the light element content of the core is likely to have been set during core formation, as was the case for the siderophile elements. As discussed earlier, studies of the metal–silicate partitioning of the MSEs indicate that core–mantle partitioning is consistent with metal–silicate equilibration at average conditions of 30–60 GPa and ≤4000 K (**Table 3**). Thus, metal–silicate partitioning of light elements (e.g., O and Si) at such conditions may have determined the light element content of the core. Rubie et al. (2004) investigated this possibility by modeling the partitioning of oxygen (actually the FeO component) between a silicate magma ocean and liquid Fe alloy during core formation as a function of magma ocean depth. Because oxygen solubility in liquid Fe increases strongly with temperature, FeO partitions increasingly into the Fe alloy as the magma ocean depth is increased beyond 1000 km, leaving the silicate depleted in FeO (**Figure 22(a)**). The results of this model are consistent with equilibration in a magma ocean 1200–2000 km deep and with the Earth's core containing 7–8 wt% oxygen.

The Earth's mantle contains ~8 wt% FeO, whereas, based on studies of Martian meteorites, the Martian mantle is considered to contain ~18 wt% FeO. This difference can be explained by oxygen partitioning during core formation so that the possibility that the two planets have similar (or even identical) bulk compositions cannot be excluded (Rubie et al., 2004). Because Mars is a much smaller planet than Earth, temperatures and pressures in a Martian magma ocean are expected to have been relatively low, so that little FeO was extracted from the mantle during core formation (**Figure 21(b)**). This model may also explain why the mass fraction of the Martian core is smaller than that of the Earth's core.

In the model of Rubie et al. (2004), the Earth's mantle starts off in a highly oxidized state and becomes more reduced as FeO partitions into the core. This forms the basis for the model of core formation under oxidizing conditions as proposed by Siebert et al. (2013). In fact, this model is only valid when there is no partitioning of silicon into the core. This is because every mole of silicon that partitions into the core releases two moles of FeO that partition primarily into the mantle (see reaction in **Section 9.03.3.2.2.3**). Thus the initially oxidized mantle cannot be reduced to its present composition when mass balance is taken into account (Rubie et al., 2015).

One advantage of modeling the metal–silicate partitioning of oxygen during core formation, in addition to the siderophile elements, is that assumptions concerning oxygen fugacity are not required. As described earlier in **Section 9.03.3.2.2.3**, this approach was extended by Rubie et al. (2011) by including all major elements when determining the compositions of equilibrated metal and silicate through a mass balance/partitioning approach. Thus, in their model of multistage core formation, the evolution of core composition could be determined as a function of mass accreted (**Figure 16(l)**). According to this model, the Earth's core contains 8 wt% Si, <1 wt% O, and 2 wt% S (but note that these predictions disagree with the arguments of Alfe et al., 2002).

9.03.4 Summary

Theoretical arguments and geochemical observations suggest that the bulk of the mass of the Earth accreted through impacts with smaller bodies (planetesimals and embryos) within about 50 Myr of solar system formation and that at least the larger impacts generated one or more large-scale, if transient, magma oceans. Although many of the impacting bodies were

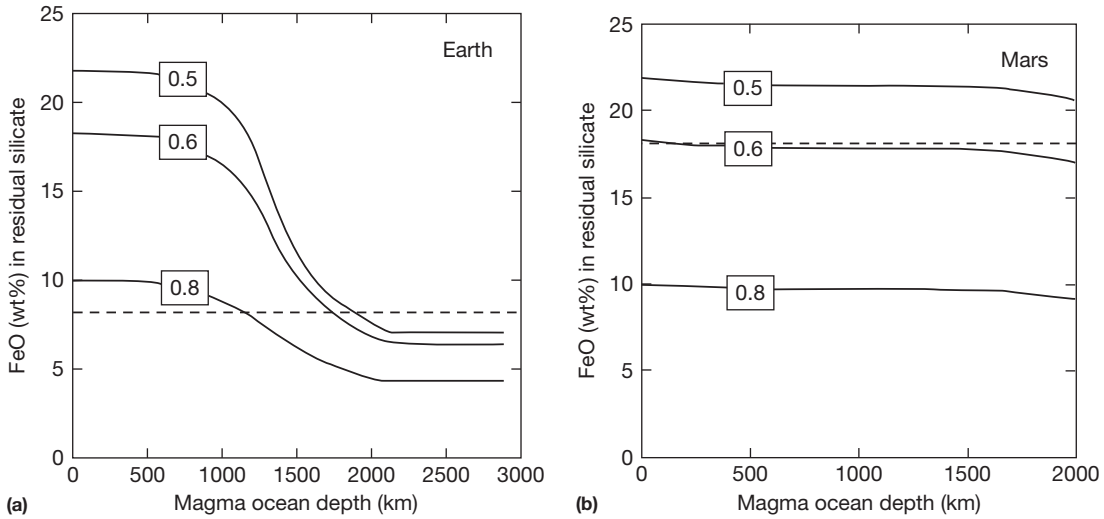


Figure 22 Results of core formation models for Earth (a) and Mars (b), based on metal–silicate partitioning of FeO in a magma ocean. The models are based on a chondritic bulk composition that consists initially of a mixture of metal and silicate components. The bulk oxygen content determines the fraction of metal that is present. The three curves, in each case, show results for different initial bulk oxygen contents, and the labels indicate the weight fraction of Fe that is initially present as metal. The horizontal dashed lines show the current FeO contents of the respective mantles. On Earth, the FeO content of the residual silicate decreases when the magma ocean depth exceeds 1000 km, because, at high temperatures, FeO partitions into the metal phase. On Mars, such an effect is almost absent because magma ocean temperatures are relatively low because of the small size of the planet. These results show that the bulk compositions of Earth and Mars could be similar (e.g., curve labeled '0.6') and that the current FeO content of the Earth's mantle resulted from core formation in a magma ocean around 1800 km deep: in this case, the Earth's core could contain 7–8 wt% oxygen. However, as discussed in the text, this model is only valid when little or no silicon partitions into the core (Reproduced from Rubie DC, Gessmann CK, and Frost DJ (2004) Partitioning of oxygen during core formation on the Earth and Mars. *Nature* 429: 58–61.).

undoubtedly differentiated, preexisting chemical signals appear to have been overprinted by the impact process. Siderophile element concentrations are consistent with magma oceans extending, *on average*, at least to midmantle depths (800–1500 km and 2500–4000 K). The impactor cores underwent partial or complete emulsification as they sank through the magma ocean, resulting in at least some degree of chemical reequilibration, as suggested by both siderophile and Hf–W isotopic observations. This reequilibration ceased as the metal pooled at the base of the magma ocean; subsequent transport of the resulting large-scale iron masses to the growing core was rapid and will have resulted in a significant increase in core temperatures. Following the Moon-forming impact, the initial core temperature was probably at least 6000 K, suggesting that extensive melting occurred in the lowermost mantle.

Unusually, there is broad agreement between the geochemical constraints and geophysical expectations of the core's early history. Nonetheless, several important outstanding questions remain as follows:

(1) The physics of exactly what happens during giant impacts is poorly understood (**Sections 9.03.2.2.2** and **9.03.2.3.2**). In particular, although there are physical and geochemical arguments for impactor emulsification, this process is still poorly understood and has not yet been fully investigated by numerical or laboratory models.
(2) The lifetime of magma oceans is also poorly known (**Section 9.03.2.3.2**). This is in part because complicating factors such as the possible presence of an insulating atmosphere or a foundering crust have a large effect on the outcome. It may be that this is an issue that can only be resolved using radiogenic isotopes with appropriate half-lives, rather than geophysical modeling. In addition, there is a possibility that terrestrial magma oceans may crystallize from the top down rather than from the bottom up (Mosenfelder et al., 2007, 2009) – which might have major implications by greatly extending magma ocean lifetimes.
(3) Geochemical models of core formation are currently hampered by a lack of partitioning data at high pressures and temperatures (i.e., >25 GPa and >3000 K). Even the partitioning of MSEs is poorly known at pressures significantly above 25 GPa and consequently large extrapolations of experimental data are required (e.g., Rubie et al., 2011). Rectifying the latter deficiency requires LH-DAC experiments (e.g., Bouhifd and Jephcoat, 2003; Tschauner et al., 1999), which are difficult to perform successfully because of large temperature gradients (which can drive chemical diffusion), temperature uncertainties, and difficulties in sample analysis. However, recent developments in this area appear very promising (Bouhifd and Jephcoat, 2011; Siebert et al., 2012).
(4) Uncertainties remain concerning the identity of the light element(s) in the core. Based on cosmochemical arguments and recent high-pressure studies, silicon with a lesser amount of oxygen may be the main light elements, together with ~2 wt% S.

(5) Many of the models of accretion and core formation to date have assumed single-stage processes. In practice, of course, accretion and core formation occur as a series of discrete events, under evolving conditions. The effect of these changing conditions on the behavior and chemistry of the core and mantle is just beginning to be addressed (e.g., Halliday, 2004; Rubie et al., 2011, 2012; Wade and Wood, 2005; Wood et al., 2006).

Acknowledgments

We thank Craig Agnor, Dan Frost, Alessandro Morbidelli, David O'Brien, Herbert Palme, Jon Wade, Bernie Wood, and Michael Walter for discussions. Portions of this work were supported by NSF EAR, NASA Origins, the German Science Foundation Priority Programme 1385 'The First 10 Million Years of the Solar System – a Planetary Materials Approach' (Grant Ru1323/2), and European Research Council (ERC) Advanced Grant 'ACCRETE' (contract no. 290568 to DCR).

References

Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.

Agnor C and Asphaug E (2004) Accretion efficiency during planetary collisions. *Astrophysical Journal* 613: L157–L160.

Agnor CB, Canup RM, and Levison HF (1999) On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* 142: 219–237.

Alfe D, Gillan MJ, and Price GD (2002) Composition and temperature of the Earth's core constrained by combining ab initio calculations and seismic data. *Earth and Planetary Science Letters* 195: 91–98.

Allegre CJ, Poirier J-P, Humler E, and Hofmann AW (1995) The chemical composition of the Earth. *Earth and Planetary Science Letters* 134: 515–526.

Anders E and Grevesse N (1989) Abundances of the elements: Meteoritic and solar. *Geochimica et Cosmochimica Acta* 53: 197–214.

Anderson OL and Isaak DG (2002) Another look at the core density deficit of Earth's outer core. *Physics of the Earth and Planetary Interiors* 131: 19–27.

Andrault D, Bolfan-Casanova N, Lo Nigro G, Bouhifd MA, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259.

Asahara Y, Frost DJ, and Rubie DC (2007) Partitioning of FeO between magnesiowüstite and liquid iron at high pressures and temperatures: Implications for the composition of the Earth's outer core. *Earth and Planetary Science Letters* 257: 435–449.

Asphaug E (2010) Similar-sized collisions and the diversity of planets. *Chemie der Erde* 70: 199219.

Asphaug E, Agnor CB, and Williams Q (2006) Hit-and-run planetary collisions. *Nature* 439: 155–160.

Badro J, Fiquet G, Guyot F, et al. (2007) Effect of light elements on the sound velocities in solid iron: Implications for the composition of Earth's core. *Earth and Planetary Science Letters* 254: 233–238.

Bagdassarov N, Golabek GJ, Solferino G, and Schmidt MW (2009a) Constraints on the Fe–S melt connectivity in mantle silicates from electrical impedance measurements. *Physics of the Earth and Planetary Interiors* 177: 139–146.

Bagdassarov N, Solferino G, Golabek GJ, and Schmidt MW (2009b) Centrifuge assisted percolation of Fe–S melts in partially molten peridotite: Time constraints for planetary core formation. *Earth and Planetary Science Letters* 288: 84–95.

Baker J, Bizzarro M, Wittig N, Connelly J, and Haack H (2005) Early planetesimal melting from an age of 4.5662 Gyr for differentiated meteorites. *Nature* 436: 1127–1131.

Ballhaus C and Ellis DJ (1996) Mobility of core melts during Earth's accretion. *Earth and Planetary Science Letters* 143: 137–145.

Benz W, Anic A, Horner J, and Whitby JA (2007) The origin of mercury. *Space Science Reviews* 132: 189202.

Benz W and Cameron AGW (1990) Terrestrial effects of the giant impact. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 61–67. Oxford: Oxford University Press.

Birch F (1952) Elasticity and constitution of the Earth's interior. *Journal of Geophysical Research* 69: 227–286.

Blichert-Toft J, Gleason JD, Télouk P, and Albaréde F (1999) The Lu–Hf isotope geochemistry of shergottites and the evolution of the Martian mantle–crust system. *Earth and Planetary Science Letters* 173: 25–39.

Boehler R (2000) High-pressure experiments and the phase diagram of lower mantle and core materials. *Reviews of Geophysics* 38: 221–245.

Borg LE and Draper DS (2003) A petrogenetic model for the origin and compositional variation of the martian basaltic meteorites. *Meteoritics and Planetary Science* 38: 1713–1731.

Bottke WF, Vokrouhlicky D, Minton D, et al. (2012) An Archaean heavy bombardment from a destabilized extension of the asteroid belt. *Nature* 485: 78–81.

Bottke WF, Walker RJ, Day JMD, Nesvorny D, and Elkins-Tanton L (2010) Stochastic late accretion to Earth, the Moon, and Mars. *Science* 330: 1527–1530.

Bouhifd MA and Jephcoat AP (2003) The effect of pressure on the partitioning of Ni and Co between silicate and iron-rich metal liquids: A diamond-anvil cell study. *Earth and Planetary Science Letters* 209: 245–255.

Bouhifd MA and Jephcoat AP (2011) Convergence of Ni and Co metal–silicate partition coefficients in the deep magma-ocean and coupled silicon–oxygen solubility in iron melts at high pressures. *Earth and Planetary Science Letters* 307: 341–348.

Boyet M and Carlson RW (2005) Nd-142 evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.

Brandon AD and Walker RJ (2005) The debate over core–mantle interaction. *Earth and Planetary Science Letters* 232: 211–225.

Brenan JM and McDonough WF (2009) Core formation and metal–silicate fractionation of osmium and iridium from gold. *Nature Geoscience* 2: 798–801.

Bruhn D, Groebner N, and Kohlstedt DL (2000) An interconnected network of core-forming melts produced by shear deformation. *Nature* 403: 883–886.

Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of Earth's outer core. *Science* 290: 1338–1342.

Buffett BA and Seagle CT (2010) Stratification of the top of the core due to chemical interactions with the mantle. *Journal of Geophysical Research* 115: B04407. http://dx.doi.org/10.1029/2009JB006751.

Busso M, Gallino R, and Wasserburg GJ (1999) Nucleosynthesis in asymptotic giant branch stars: Relevance for galactic enrichment and solar system formation. *Annual Review of Astronomy and Astrophysics* 37: 239–309.

Cameron AGW (2000) High-resolution simulations of the giant impact. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 133–144. Tucson: University of Arizona Press.

Canup RM (2004) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.

Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.

Canup RM and Righter K (eds.) (2000) *Origin of the Earth and Moon*. Tucson: University of Arizona Press.

Canup RM, Ward WR, and Cameron AGW (2001) A scaling relationship for satellite-forming impacts. *Icarus* 150: 288–296.

Caro G, Bourdon B, Halliday AN, and Quitte G (2008) Super-chondritic Sm/Nd ratios in Mars, the Earth and the Moon. *Nature* 452: 336–339.

Caro G, Bourdon B, Wood BJ, and Corgne A (2005) Trace-element fractionation in Hadean mantle generated by melt segregation from a magma ocean. *Nature* 436: 246–249.

Chabot NL and Agee CB (2003) Core formation in the Earth and Moon: New experimental constraints from V, Cr, and Mn. *Geochimica et Cosmochimica Acta* 67: 2077–2091.

Chabot NL, Draper DS, and Agee CB (2005) Conditions of core formation in the Earth: Constraints from nickel and cobalt partitioning. *Geochimica et Cosmochimica Acta* 69: 2141–2151.

Chabot NL and Haack H (2006) Evolution of asteroidal cores. In: Lauretta DS and McSween HY (eds.) *Meteorites and the Early Solar System*, pp. 747–771. Tucson, AZ: University of Arizona Press.

Chamberlin TC (1916) *The Origin of the Earth*, p. 271. Chicago: University of Chicago Press.

Chambers JE (2008) Oligarchic growth with migration and fragmentation. *Icarus* 198: 256–273.

Chambers JE (2009) Planetary migration: What does it mean for planet formation? *Annual Review of Earth and Planetary Sciences* 37: 321–344.

Chambers JE (2010) Terrestrial planet formation. In: Seager S (ed.) *Exoplanets*, pp. 297–318. Tucson: University of Arizona Press.

Chambers JE and Wetherill GW (1998) Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus* 136: 304–327.

Clayton RN and Mayeda TK (1996) Oxygen isotope studies of achondrites. *Geochimica et Cosmochimica Acta* 60: 1999–2017.

Corgne A, Siebert J, and Badro J (2009) Oxygen as a light element: A solution to single-stage core formation. *Earth and Planetary Science Letters* 288: 108–114.

Cottrell E and Walker D (2006) Constraints on core formation from Pt partitioning in mafic silicate liquids at high temperatures. *Geochimica et Cosmochimica Acta* 70: 1565–1580.
Cottrell E, Walter MJ, and Walker D (2009) Metal–silicate partitioning of tungsten at high pressure and temperature: Implications for equilibrium core formation in Earth. *Earth and Planetary Science Letters* 281: 275–287.
Dahl TW and Stevenson DJ (2010) Turbulent mixing of metal and silicate during planet accretion—And interpretation of the Hf–W chronometer. *Earth and Planetary Science Letters* 295: 177–186.
Dalziel SB, Linden PF, and Youngs DL (1999) Self-similarity and internal structure of turbulence induced by Rayleigh–Taylor instability. *Journal of Fluid Mechanics* 399: 1–48.
Dauphas N and Pourmand A (2011) Hf–W–Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* 473: 489–492.
Day JMD, Walker RJ, Qin L, et al. (2012) Late accretion as a natural consequence of planetary growth. *Nature Geoscience* 5: 614–617.
Debaille V, Brandon AD, Yin QZ, and Jacobsen B (2007) Coupled ^{142}Nd–^{143}Nd evidence for a protracted magma ocean in Mars. *Nature* 450: 525–528.
Deguen R, Olson P, and Cardin P (2011) Experiments on turbulent metal–silicate mixing in a magma ocean. *Earth and Planetary Science Letters* 310: 303–313.
Drake MJ (2001) The eucrite/Vesta story. *Meteoritics and Planetary Science* 36: 501–513.
Drake MJ and Righter K (2002) Determining the composition of the Earth. *Nature* 416: 39–44.
Dreibus G and Palme H (1996) Cosmochemical constraints on the sulfur content in the Earth's core. *Geochimica et Cosmochimica Acta* 60: 1125–1130.
Dwyer CA, Nimmo F, and Chambers JE (2015) Bulk chemical and Hf-W isotopic consequences of incomplete accretion during planet formation. *Icarus* 245: 145–152.
Elkins-Tanton LT, Burgess S, and Yin Q-Z (2011) The lunar magma ocean: Reconciling the solidification process with lunar petrology and geochemistry. *Earth and Planetary Science Letters* 304: 326–336.
Elkins-Tanton LT, Parmentier EM, and Hess PC (2003) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38: 1753–1771.
Elsasser WM (1963) Early history of the earth. In: Geiss J and Goldberg ED (eds.) *Earth Science and Meteoritics*, pp. 1–30. Amsterdam: North-Holland.
Elser S, Meyer MR, and Moore B (2012) On the origin of elemental abundances in the terrestrial planets. *Icarus* 221: 859–874.
Ertel W, O'Neill HStC, Sylvester PJ, and Dingwell DB (1999) Solubilities of Pt and Rh in a haplobasaltic silicate melt at 1300 °C. *Geochimica et Cosmochimica Acta* 63: 2439–2449.
Ertel W, O'Neill HStC, Sylvester PJ, Dingwell DB, and Spettel B (2001) The solubility of rhenium in silicate melts. Implications for the geochemical properties of rhenium at high temperatures. *Geochimica et Cosmochimica Acta* 65: 2161–2170.
Ertel W, Walter MJ, Drake MJ, and Sylvester PJ (2006) Experimental study of platinum solubility in silicate melt to 14 GPa and 2273 K: Implications for accretion and core formation in the Earth. *Geochimica et Cosmochimica Acta* 70: 2591–2602.
Faul UH (1997) Permeability of partially molten upper mantle rocks from experiments and percolation theory. *Journal of Geophysical Research* 102: 10299–10312. http://dx.doi.org/10.1029/96JB03460.
Fiquet G, Auzende AL, Siebert J, et al. (2010) Melting of peridotite to 140 GPa. *Science* 329: 1516–1518.
Fitoussi C, Bourdon B, Kleine T, Oberli F, and Reynolds BC (2009) Si isotope systematics of meteorites and terrestrial peridotites: Implications for Mg/Si fractionation in the solar nebula and for Si in the Earth's core. *Earth and Planetary Science Letters* 287: 77–85.
Fortenfant SS, Dingwell DB, Ertel-Ingrisch W, Capmas F, Birck JL, and Dalpé C (2006) Oxygen fugacity dependence of Os solubility in haplobasaltic melt. *Geochimica et Cosmochimica Acta* 70: 742–756.
Fortenfant SS, Günther D, Dingwell DB, and Rubie DC (2003a) Temperature dependence of Pt and Rh solubilities in a haplobasaltic melt. *Geochimica et Cosmochimica Acta* 67: 123–131.
Fortenfant SS, Rubie DC, Reid J, Dalpé C, and Gessmann CK (2003b) Partitioning of Re and Os between liquid metal and magnesiowüstite at high pressure. *Physics of the Earth and Planetary Interiors* 139: 77–91.
Frost DJ, Asahara Y, Rubie DC, et al. (2010) The partitioning of oxygen between the Earth's mantle and core. *Journal of Geophysical Research* 115: B02202. http://dx.doi.org/10.1029/2009JB006302.
Frost DJ, Liebske C, Langenhorst F, McCammon CA, Tronnes RG, and Rubie DC (2004) Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle. *Nature* 428: 409–412.
Frost DJ, Mann U, Asahara Y, and Rubie DC (2008) The redox state of the mantle during and just after core formation. *Philosophical Transactions of the Royal Society A* 366: 4315–4337. http://dx.doi.org/10.1098/rsta.2008.0147.
Fu RR, Weiss BP, Shuster DL, et al. (2012) An ancient core dynamo in asteroid Vesta. *Science* 338: 238–241.
Gaetani GA and Grove TL (1999) Wetting of mantle olivine by sulfide melt: Implications for Re/Os ratios in mantle peridotite and late-stage core formation. *Earth and Planetary Science Letters* 169: 147–163.
Gancarz AJ and Wasserburg GJ (1977) Initial Pb of the Amîtsoq gneiss, West Greenland, and implications for the age of the Earth. *Geochim. Cosmochim. Acta* 41(9): 1283–1301.
Georg RB, Halliday AN, Schauble EA, and Reynolds BC (2007) Silicon in the Earth's core. *Nature* 447: 1102–1106. http://dx.doi.org/10.1038/nature05927.
Gessmann CK and Rubie DC (1998) The effect of temperature on the partitioning of Ni, Co, Mn, Cr and V at 9 GPa and constraints on formation of the Earth's core. *Geochimica et Cosmochimica Acta* 62: 867–882.
Gessmann CK and Rubie DC (2000) The origin of the depletions of V, Cr and Mn in the mantles of the Earth and Moon. *Earth and Planetary Science Letters* 184: 95–107.
Gessmann CK, Wood BJ, Rubie DC, and Kilburn MR (2001) Solubility of silicon in liquid metal at high pressure: Implications for the composition of the Earth's core. *Earth and Planetary Science Letters* 184: 367–376.
Ghosh A and McSween HY (1998) A thermal model for the differentiation of Asteroid 4 Vesta, based on radiogenic heating. *Icarus* 134: 187–206.
Golabek GJ, Schmeling H, and Tackley PJ (2008) Earth's core formation aided by flow channeling instabilities induced by iron diapirs. *Earth and Planetary Science Letters* 271: 24–33.
Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.
Goossens S, et al. (2011) Lunar gravity field determination using Selene same-beam differential VLBI tracking data. *Journal of Geodesy* 85: 205–228.
Greenwood RC, Franchi IA, Jambon A, and Buchanan PC (2005) Widespread magma oceans on asteroid bodies in the early solar system. *Nature* 435: 916–918. http://dx.doi.org/10.1038/nature03612.
Grimm RE and McSween HY (1993) Heliocentric zoning of the asteroid belt by aluminum-26 heating. *Science* 259: 653–655.
Groebner N and Kohlstedt DL (2006) Deformation-induced metal melt networks in silicates: Implications for core–mantle interactions in planetary bodies. *Earth and Planetary Science Letters* 245: 571–580.
Haisch KE, Lada EA, and Lada CJ (2001) Disk frequencies and lifetimes in young clusters. *Astrophysical Journal* 553: L153–L156.
Halliday AN (2003) The origin and earliest history of the Earth. *Treatise on Geochemistry*, vol. 1, pp. 509–557. Oxford: Elsevier.
Halliday AN (2004) Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* 427: 505–509.
Hallworth MA, Phillips JC, Hubbert HE, and Sparks RSJ (1993) Entrainment in turbulent gravity currents. *Nature* 362: 829–831.
Hanks TC and Anderson DL (1969) The early thermal history of the Earth. *Physics of the Earth and Planetary Interiors* 2: 19–29.
Helffrich G and Kaneshima S (2004) Seismological constraints on core composition from Fe–O–S liquid immiscibility. *Science* 306: 2239–2242.
Hillgren VJ, Drake MJ, and Rubie DC (1994) High-pressure and high-temperature experiments on core–mantle segregation in the accreting Earth. *Science* 264: 1442–1445.
Hillgren VJ, Gessmann CK, and Li J (2000) An experimental perspective on the light element in Earth's core. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 245–264. Tucson: University of Arizona Press.
Höink T, Schmalzl J, and Hansen U (2006) Dynamics of metal–silicate separation in a terrestrial magma ocean. *Geochemistry, Geophysics, Geosystems* 7: Q09008. http://dx.doi.org/10.1029/2006GC001268.
Holzheid A, Schmitz MD, and Grove TL (2000a) Textural equilibria of iron sulphide liquids in partly molten silicate aggregates and their relevance to core formation scenarios. *Journal of Geophysical Research* 105: 13555–13567.
Holzheid A, Sylvester P, O'Neill HStC, Rubie DC, and Palme H (2000b) Evidence for a late chondritic veneer in the Earth's mantle from high-pressure partitioning of palladium and platinum. *Nature* 406: 396–399.
Hustoft JW and Kohlstedt DL (2006) Metal–silicate segregation in deforming dunitic rocks. *Geochemistry, Geophysics, Geosystems* 7: Q02001. http://dx.doi.org/10.1029/2005GC001048.
Iida T and Guthrie RL (1988) *The Physical Properties of Liquid Metals*, pp. 109–146. Oxford: Clarendon.
Jacobsen SB (2005) The Hf–W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Science* 33: 531–570.

Javoy M (1995) The integral enstatite chondrite model of the Earth. *Geophysical Research Letters* 22: 2219–2222.

Jeffreys H (1952) *The Earth*, 3rd ed. Cambridge: Cambridge University Press.

Jephcoat A and Olson P (1987) Is the inner core of the Earth pure iron? *Nature* 325: 332–335.

Jones JH, Capobianco CJ, and Drake MJ (1992) Siderophile elements and the Earth's formation. *Science* 257: 1281–1282.

Jones JH, Neal CR, and Ely JC (2003) Signatures of the siderophile elements in the SNC meteorites and Mars: A review and petrologic synthesis. *Chemical Geology* 196: 21–41.

Kanda R and Stevenson DJ (2006) Suction mechanism for iron entrainment into the lower mantle. *Geophysical Research Letters* 33: L02310. http://dx.doi.org/10.1029/2005GL025009.

Karato S-I and Murthy VR (1997) Core formation and chemical equilibrium in the Earth—I. Physical considerations. *Physics of the Earth and Planetary Interiors* 100: 61–79.

Karki BB and Stixrude LP (2010) Viscosity of $MgSiO_3$ liquid at Earth's mantle conditions: Implications for an early magma ocean. *Science* 328: 740–742.

Kato T and Ringwood AE (1989) Melting relationships in the system Fe–FeO at high pressure: Implications for the composition and formation of the Earth's core. *Physics and Chemistry of Minerals* 16: 524–538.

Kaula WM (1979) Thermal evolution of the Earth and Moon growing by planetesimal impacts. *Journal of Geophysical Research* 84: 999–1008.

Kegler P, Holzheid A, Frost DJ, Rubie DC, Dohmen R, and Palme H (2008) New Ni and Co metal–silicate partitioning data and their relevance for an early terrestrial magma ocean. *Earth and Planetary Science Letters* 268: 28–40.

Kegler P, Holzheid A, Rubie DC, Frost DJ, and Palme H (2005) New results of metal/silicate partitioning of Ni and Co at elevated pressures and temperatures. In: *XXXVI Lunar and Planetary Science Conference,* Abstr. #2030.

Keil K, Stoeffler D, Love SG, and Scott ERD (1997) Constraints on the role of impact heating and melting in asteroids. *Meteoritics and Planetary Science* 32: 349–363.

Kendall JD and Melosh HJ (2012a) Dispersion of iron cores during planetesimal impacts. In: *Meteoritics and Planetary Science Supplement,* Abstract 5302.

Kendall JD and Melosh HJ (2012b) Fate of iron cores during planetesimal impacts. In: *LPSC Conference, 43,* Abstract #2699.

Keppler H and Rubie DC (1993) Pressure-induced coordination changes of transition-metal ions in silicate melts. *Nature* 364: 54–55.

Kilburn MR and Wood BJ (1997) Metal–silicate partitioning and the incompatibility of S and Si during core formation. *Earth and Planetary Science Letters* 152: 139–148.

Kimura K, Lewis RS, and Anders E (1974) Distribution of gold between nickel–iron and silicate melts: Implications for the abundances of siderophile elements on the Earth and Moon. *Geochimica et Cosmochimica Acta* 38: 683–701.

Kleine T, Palme H, Mezger M, and Halliday AN (2005) Hf–W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* 310: 1671–1673.

Kleine T, Touboul M, Bourdon B, et al. (2009) Hf–W chronometry and the accretion and early evolution of asteroids and terrestrial planets. *Geochimica et Cosmochimica Acta* 73: 5150–5188.

Kokubo E and Genda H (2010) Formation of terrestrial planets from protoplanets under a realistic accretion condition. *Astrophysical Journal Letters* 714: L21–L25.

Kokubo E and Ida S (1998) Oligarchic growth of protoplanets. *Icarus* 131: 171–178.

Kominami J, Tanaka H, and Ida S (2005) Orbital evolution and accretion of protoplanets tidally interacting with a gas disk I. Effects of interaction with planetesimals and other protoplanets. *Icarus* 178: 540–552.

Kong P, Ebihara M, and Palme H (1999) Siderophile elements in Martian meteorites and implications for core formation in Mars. *Geochimica et Cosmochimica Acta* 63: 1865–1875.

Kyte FT (1998) A meteorite from the Cretaceous/Tertiary boundary. *Nature* 396: 237–239.

Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450: 866–869. http://dx.doi.org/10.1038/nature06355.

Laporte D and Watson EB (1995) Experimental and theoretical constraints on melt distribution in crustal sources: The effect of crystalline anisotropy on melt interconnectivity. *Chemical Geology* 124: 161–184.

Leinhardt ZM and Richardson DC (2005) Planetesimals to protoplanets 1. Effect of fragmentation on terrestrial planet formation. *Astrophysical Journal* 625: 427–440.

Li J and Agee CB (1996) Geochemistry of mantle–core differentiation at high pressure. *Nature* 381: 686–689.

Li J and Agee CB (2001) The effect of pressure, temperature, oxygen fugacity and composition on partitioning of nickel and cobalt between liquid Fe–Ni–S alloy and liquid silicate: Implications for the Earth's core formation. *Geochimica et Cosmochimica Acta* 65: 1821–1832.

Li J and Fei Y (2003) Experimental constraints on core composition. In: Holland HD and Turekian KK (eds.) Carlson RW (ed.) *The Mantle and Core*. *Treatise on Geochemistry*, vol. 2, pp. 521–546. Oxford: Elsevier-Pergamon.

Liebske C and Frost DJ (2012) Melting phase relations in the MgO–$MgSiO_3$ system between 16 and 26 GPa: Implications for melting in Earth's deep interior. *Earth and Planetary Science Letters* 345–348: 159–170.

Liebske C, Schmickler B, Terasai H, et al. (2005) Viscosity of peridotite liquid up to 13 GPa: Implications for magma ocean viscosities. *Earth and Planetary Science Letters* 240: 589–604.

Lin J-R, Gerya TV, Tackley PJ, Yuen DA, and Golabek GJ (2009) Numerical modeling of protocore destabilization during planetary accretion: Methodology and results. *Icarus* 204: 732–748.

Lin J-R, Gerya TV, Tackley PJ, Yuen DA, and Golabek GJ (2011) Protocore destabilization in planetary embryos formed by cold accretion: Feedbacks from non-Newtonian rheology and energy dissipation. *Icarus* 213: 24–42.

Lindstrom DJ and Jones JH (1996) Neutron activation analysis of multiple 10–100 μg glass samples from siderophile element partitioning experiments. *Geochimica et Cosmochimica Acta* 60: 1195–1203.

Lodders K (2003) Solar System abundances and condensation temperatures of the elements. *The Astrophysical Journal* 591: 1220–1247.

Lodders K and Fegley B (1998) *The Planetary Scientist's Companion*. Oxford: Oxford University Press.

Ma Z (2001) Thermodynamic description for concentrated metallic solutions using interaction parameters. *Metallurgical and Materials Transactions B* 32: 87–103.

Maier WD, Andreoli MAG, McDonald I, et al. (2006) Discovery of a 25 cm asteroid clast in the giant Morokweng impact crater, South Africa. *Nature* 441: 203–206.

Malavergne V, Siebert J, Guyot F, et al. (2004) Si in the core? New high-pressure and high-temperature experimental data. *Geochimica et Cosmochimica Acta* 68: 4201–4211.

Mann U, Frost DJ, and Rubie DC (2009) Evidence for high-pressure core–mantle differentiation from the metal–silicate partitioning of lithophile and weakly-siderophile elements. *Geochimica et Cosmochimica Acta* 73: 7360–7386.

Mann U, Frost DJ, Rubie DC, Becker H, and Audétat A (2012) Partitioning of Ru, Rh, Pd, Re, Ir and Pt between liquid metal and silicate at high pressures and high temperatures – Implications for the origin of highly siderophile element concentrations in the Earth's mantle. *Geochimica et Cosmochimica Acta* 84: 593–613. http://dx.doi.org/10.1016/j.gca.2012.01.026.

Margot JL, Peale SJ, and Solomon SC (2012) Mercury's moment of inertia from spin and gravity data. *Journal of Geophysical Research* 117: E00109.

Matsui T and Abe Y (1986) Impact-induced atmospheres and oceans on Earth and Venus. *Nature* 322: 526–528.

McDonough WF (2003) Compositional model for the Earth's core. *Treatise on Geochemistry,* pp. 517–568. Amsterdam: Elsevier.

McDonough WF and Sun S-s (1995) The composition of the Earth. *Chemical Geology* 120: 223–253.

McSween HY (1999) *Meteorites and Their Parent Planets*. Cambridge, UK: Cambridge University Press.

Médard E, Schmidt MW, Wälle M, Keller NS, and Günther D (2010) Pt in silicate melts: Centrifuging nanonuggets to decipher core formation processes. *41st Lunar and Planetary Science Conference,* Abstract #2639.

Melosh HJ (1990) Giant impacts and the thermal state of the early Earth. In: Newsom HE and Jones JE (eds.) *Origin of the Earth*, pp. 69–83. Oxford: Oxford University Press.

Melosh HJ and Rubie DC (2007) Ni partitioning in the terrestrial magma ocean: A polybaric numerical model. In: *Lunar and Planetary Science XXXVIII,* Abstract #1593.

Miller GH, Stolper EM, and Ahrens TJ (1991) The equation of state of a molten komatiite 2. Application to komatiite petrogenesis and the Hadean mantle. *Journal of Geophysical Research* 96: 11849–11864.

Minarik WG, Ryerson FJ, and Watson EB (1996) Textural entrapment of core-forming melts. *Science* 272: 530–533.

Monteux J, Ricard Y, Coltice N, Dubuffet F, and Ulvrova M (2009) A model of metal–silicate separation on growing planets. *Earth and Planetary Science Letters* 287: 353–362.

Morbidelli A, Chambers J, Lunine JI, et al. (2000) Source regions and timescales for the delivery of water to the Earth. *Meteoritics and Planetary Science* 35: 1309–1320.

Morgan JW (1986) Ultramafic xenoliths: Clues to Earth's late accretionary history. *Journal of Geophysical Research* 91: 12375–12387.

Morishima R, Stadel J, and Moore B (2010) From planetesimals to terrestrial planets: N-body simulations including the effects of nebular gas and giant planets. *Icarus* 207: 517–535.

Mosenfelder JL, Asimow PD, and Ahrens TJ (2007) Thermodynamic properties of Mg_2SiO_4 liquid at ultra-high pressures from shock measurements to 200 GPa on

forsterite and wadsleyite. *Journal of Geophysical Research* 112: B06208. http://dx.doi.org/10.1029/2006JB004364.
Mosenfelder JL, Asimow PD, Frost DJ, Rubie DC, and Ahrens TJ (2009) The $MgSiO_3$ system at high pressure: Thermodynamic properties of perovskite, postperovskite, and melt from global inversion of shock and static compression data. *Journal of Geophysical Research* 114: B01203. http://dx.doi.org/10.1029/2008JB005900.
Mukhopadhyay S (2012) Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* 486: 101–124.
Murthy VR (1991) Early differentiation of the Earth and the problem of mantle siderophile elements – A new approach. *Science* 253: 303–306.
Newsom HE (1990) Accretion and core formation in the Earth: Evidence from siderophile elements. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 273–288. New York: Oxford University Press.
Nimmo F, O'Brien DP, and Kleine T (2010) Tungsten isotopic evolution during late-stage accretion: Constraints on Earth–Moon equilibration. *Earth and Planetary Science Letters* 292: 363–370.
O'Neill HStC (1991) The origin of the Moon and the early history of the Earth – A chemical model. Part 2: The Earth. *Geochimica et Cosmochimica Acta* 55: 1159–1172.
O'Neill HStC (1992) Siderophile elements and the Earth's formation. *Science* 257: 1282–1284.
O'Neill HStC and Palme H (1998) Composition of the silicate Earth: Implications for accretion and core formation. In: Jackson I (ed.) *The Earth's Mantle*, pp. 3–126. Cambridge, UK: Cambridge University Press.
O'Brien DP, Morbidelli A, and Levison HF (2006) Terrestrial planet formation with strong dynamical friction. *Icarus* 184: 39–58.
Ohtani E and Ringwood AE (1984) Composition of the core. I. Solubility of oxygen in molten iron at high temperatures. *Earth and Planetary Science Letters* 71: 85–93.
Ohtani E, Ringwood AE, and Hibberson W (1984) Composition of the core. II. Effect of high pressure on solubility of FeO in molten iron. *Earth and Planetary Science Letters* 71: 94–103.
Ohtani E and Yurimoto H (1996) Element partitioning between metallic liquid, magnesiowüstite, and silicate liquid at 20 GPa and 2500 °C: A secondary ion mass spectrometric study. *Geophysical Research Letters* 23: 1993–1996.
O'Neill HSC, Canil D, and Rubie DC (1998) Oxide–metal equilibria to 2500 °C and 25 GPa: Implications for core formation and the light component in the Earth's core. *Journal of Geophysical Research* 103: 12239–12260.
O'Neill HSC and Palme H (2008) Collisional erosion and the non-chondritic composition of the terrestrial planets. *Philosophical Transactions of the Royal Society of London A* 366: 4205–4238.
Palme H and O'Neill HStC (2003) Cosmochemical estimates of mantle composition. In: Carlson RW (ed.) and Holland HD and Turekian KK (series eds.) *The Mantle and Core. Treatise on Geochemistry*, vol. 2, pp. 1–38. Oxford: Elsevier-Pergamon.
Pierazzo E, Vickery AM, and Melosh HJ (1997) A re-evaluation of impact melt production. *Icarus* 127: 408–432.
Poirier J-P (1994) Light elements in the Earth's outer core: A critical review. *Physics of the Earth and Planetary Interiors* 85: 319–337.
Poirier J-P, Malavergne V, and Le Mouël JL (1998) Is there a thin electrically conducting layer at the base of the mantle? In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region*. *Geodynamics*, vol. 28, pp. 131–137. Washington, DC: American Geophysical Union.
Porcelli D, Woolum D, and Cassen P (2001) Deep Earth rare gases: Initial inventories, capture from the solar nebula, and losses during Moon formation. *Earth and Planetary Science Letters* 193: 237–251.
Pritchard ME and Stevenson DJ (2000) Thermal aspects of a lunar origin by giant impact. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tucson: University of Arizona Press.
Quitté G, Latkoczy C, Halliday AN, Schönbächler M, and Günther D (2005) Iron-60 in the Eucrite parent body and the initial $^{60}Fe/^{56}Fe$ of the solar system. In: *LPSC-XXXVI*, Abstract 1827.
Raymond SN, Quinn T, and Lunine JI (2007) High resolution simulations of the final assembly of Earth-like planets 2. Water delivery and planetary habitability. *Astrobiology* 7: 66–84.
Reid JE, Suzuki A, Funakoshi K, et al. (2003) The viscosity of $CaMgSi_2O_6$ liquid at pressures up to 13 GPa. *Physics of the Earth and Planetary Interiors* 139: 45–54.
Righter K (2003) Metal–silicate partitioning of siderophile elements and core formation in the early Earth. *Annual Review of Earth and Planetary Sciences* 31: 135–174.
Righter K (2005) Highly siderophile elements: Constraints on Earth accretion and early differentiation. In: van der Hilst RD, Bass JD, Matas J, and Trampert J (eds.) *Earth's Deep Mantle: Structure, Composition and Evolution*. *Geophysical Monograph*, vol. 160, pp. 201–218. Washington: American Geophysical Union.
Righter K (2011) Prediction of metal–silicate partition coefficients for siderophile elements: An update and assessment of PT conditions for metal–silicate equilibrium during accretion of the Earth. *Earth and Planetary Science Letters* 304: 158–167.
Righter K and Drake MJ (1996) Core formation in Earth's Moon, Mars, and Vesta. *Icarus* 124: 513–529.
Righter K and Drake MJ (1997) Metal–silicate equilibrium in a homogeneously accreting earth: New results for Re. *Earth and Planetary Science Letters* 146: 541–553.
Righter K and Drake MJ (1999) Effect of water on metal–silicate partitioning of siderophile elements: A high pressure and temperature terrestrial magma ocean and core formation. *Earth and Planetary Science Letters* 171: 383–399.
Righter K and Drake MJ (2000) Metal–silicate equilibration in the early Earth: New constraints from volatile moderately siderophile elements Ga, Sn, Cu and P. *Geochimica et Cosmochimica Acta* 64: 3581–3597.
Righter K and Drake MJ (2001) Constraints on the depth of an early terrestrial magma ocean. *Meteoritics & Planetary Science* 36: A173.
Righter K and Drake MJ (2003) Partition coefficients at high pressure and temperature. *Treatise on Geochemistry*, vol. 2, pp. 425–449. Amsterdam: Elsevier.
Righter K, Drake MJ, and Yaxley G (1997) Prediction of siderophile element metal–silicate partition coefficients to 20 GPa and 2800 °C: The effects of pressure, temperature, oxygen fugacity, and silicate and metallic melt compositions. *Physics of the Earth and Planetary Interiors* 100: 115–134.
Righter K, Humayun M, and Danielson L (2008) Partitioning of palladium at high pressures and temperatures during core formation. *Nature Geoscience* 1: 321–323.
Righter K, Pando KM, Danielson L, and Lee C-T (2010) Partitioning of Mo, P and other siderophile elements (Cu, Ga, Sn, Ni, Co, Cr, Mn, V, and W) between metal and silicate melt as a function of temperature and silicate melt composition. *Earth and Planetary Science Letters* 291: 1–9.
Ringwood AE (1977) Composition of the core and implications for the origin of the Earth. *Geochemical Journal* 11: 111–135.
Rose LA and Brenan JM (2001) Wetting properties of Fe–Ni–Co–Cu–O–S melts against olivine: Implications for sulfide melt mobility. *Economic Geology and the Bulletin of the Society of Economic Geologists* 96: 145–157.
Rubie DC, Jacobson SA, Morbidelli A, et al. (2015) Accretion and differentiation of the terrestrial planets with implications for the compositions of early-formed Solar System bodies and accretion of water. *Icarus* 248: 89–108.
Rubie DC, Frost DJ, Mann U, et al. (2011) Heterogeneous accretion, composition and core–mantle differentiation of the Earth. *Earth and Planetary Science Letters* 301: 31–42. http://dx.doi.org/10.1016/j.epsl.2010.11.030.
Rubie DC, Gessmann CK, and Frost DJ (2004) Partitioning of oxygen during core formation on the Earth and Mars. *Nature* 429: 58–61.
Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.
Rubin AM (1993) Dikes vs. diapirs in viscoelastic rock. *Earth and Planetary Science Letters* 119: 641–659.
Rubin AM (1995) Propagation of magma-filled cracks. *Annual Review of Earth and Planetary Sciences* 23: 287–336.
Rudge JF, Kleine T, and Bourdon B (2010) Broad bounds on Earth's accretion and core formation constrained by geochemical models. *Nature Geoscience* 3: 439–443.
Rushmer T, Minarik WG, and Taylor GJ (2000) Physical processes of core formation. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 227–244. Tucson: University of Arizona Press.
Rushmer T, Petford N, Humayun M, and Campbell AJ (2005) Fe-liquid segregation in deforming planetesimals: Coupling core-forming compositions with transport phenomena. *Earth and Planetary Science Letters* 239: 185–202.
Russell CT and Luhmann JG (1997) Mercury: Magnetic field and magnetosphere. In: Shirley JH and Fairbridge RW (eds.) *Encyclopedia of Planetary Sciences*, pp. 476–478. New York: Chapman and Hall.
Safronov VS (1978) The heating of the Earth during its formation. *Icarus* 33: 1–12.
Sakai T, Kanto T, Ohtani E, et al. (2006) Interaction between iron and post-perovskite at core–mantle boundary and core signature in plume source region. *Geophysical Research Letters* 33: L15317. http://dx.doi.org/10.1029/2006GL026868.
Samuel H (2012) A re-evaluation of metal diapir breakup and equilibration in terrestrial magma oceans. *Earth and Planetary Science Letters* 313–314: 105–114.
Samuel H, Tackley PJ, and Evonuk M (2010) Heat partitioning in terrestrial planets during core formation by negative diapirism. *Earth and Planetary Science Letters* 290: 13–19.

Sanloup C, van Westrenen W, Dasgupta R, Maynard-Casely H, and Perrillat J-P (2011) Compressibility change in iron-rich melt and implications for core formation models. *Earth and Planetary Science Letters* 306: 118–121.

Scherstén A, Elliott T, Hawkesworth C, Russell S, and Massarik J (2006) Hf–W evidence for rapid differentiation of iron meteorite parent bodies. *Earth and Planetary Science Letters* 241: 530–542.

Schmickler B, Liebske C, Holzapfel C, and Rubie DC (2005) Viscosity of peridotite liquid up to 24 GPa: Predictions from self-diffusion coefficients. *Fall Meeting Suppl., Abstr. MR11A-06*, *EOS Transactions American Geophysical Union* 86(52).

Schönbächler M, Carlson RW, Horan MF, Mock TD, and Hauri EH (2010) Heterogeneous accretion and the moderately volatile element budget of the Earth. *Science* 328: 884–887.

Schramm DN, Tera F, and Wasserburg GJ (1970) The isotopic abundance of ^{26}Mg and limits on ^{26}Al in the early solar system. *Earth and Planetary Science Letters* 10: 44–59.

Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure and dynamics of the Galilean satellites. In: Bagenal F, Dowling T, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 281–306. Cambridge: Cambridge University Press.

Shannon MC and Agee CB (1996) High pressure constraints on percolative core formation. *Geophysical Research Letters* 23: 2717–2720.

Shannon MC and Agee CB (1998) Percolation of core melts at lower mantle conditions. *Science* 280: 1059–1061.

Siebert J, Badro J, Antonangeli D, and Ryerson FJ (2012) Metal–silicate partitioning of Ni and Co in a deep magma ocean. *Earth and Planetary Science Letters* 321–322: 189–197.

Siebert J, Badro J, Antonangeli D, and Ryerson FJ (2013) Terrestrial accretion under oxidizing conditions. *Science* 339: 1194–1197.

Siebert J, Corgne A, and Ryerson FJ (2011) Systematics of metal–silicate partitioning for many siderophile elements applied to Earth's core formation. *Geochimica et Cosmochimica Acta* 75: 1451–1489.

Snow JE and Schmidt G (1998) Constraints on Earth accretion deduced from noble metals in the oceanic mantle. *Nature* 391: 166–169.

Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tucson: University of Arizona Press.

Solomon S (1979) Formation, history and energetics of cores in the terrestrial planets. *Physics of the Earth and Planetary Interiors* 19: 168–182.

Spohn T and Schubert G (1991) Thermal equilibration of the Earth following a giant impact. *Geophysical Journal International* 107: 163–170.

Stevenson DJ (1981) Models of the Earth's core. *Science* 214: 611–619.

Stevenson DJ (1989) Formation and early evolution of the Earth. In: Peltier WR (ed.) *Mantle Convection Plate Tectonics and Global Dynamics*. New York: Gordon and Breach.

Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*. Oxford: Oxford University Press.

Stevenson DJ (2003) Mission to the Earth's core—A modest proposal. *Nature* 423: 239–240.

Stixrude L and Karki B (2005) Structure and freezing of $MgSiO_3$ liquid in Earth's lower mantle. *Science* 310: 297–299.

Tachibana S, Huss GR, Kita NT, Shimoda G, and Morishita Y (2006) Fe-60 in chondrites: Debris from a nearby supernova in the early solar system? *Astrophysical Journal* 639: L87–L90.

Takafuji N, Hirose K, Mitome M, and Bando Y (2005) Solubilities of O and Si in liquid iron in equilibrium with $(Mg,Fe)SiO_3$ perovskite and the light elements in the core. *Geophysical Research Letters* 32: L06313. http://dx.doi.org/10.1029/2005GL022773.

Takafuji N, Hirose K, Ono S, Xu F, Mitome M, and Bando Y (2004) Segregation of core melts by permeable flow in the lower mantle. *Earth and Planetary Science Letters* 224: 249–257.

Tang H and Dauphas N (2012) Abundance, distribution, and origin of 60Fe in the solar protoplanetary disk. *Earth and Planetary Science Letters* 359–360: 248–263.

Tarduno JA, Cottrell RD, Nimmo F, et al. (2012) Evidence for a dynamo in the main group pallasite parent body. *Science* 338: 93–95.

Taylor SR and Norman MD (1990) Accretion of differentiated planetesimals to the Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 29–43. New York: Oxford University Press.

Terasaki H, Frost DJ, Rubie DC, and Langenhorst F (2005) The effect of oxygen and sulphur on the dihedral angle between Fe–O–S melt and silicate minerals at high pressure: Implications for Martian core formation. *Earth and Planetary Science Letters* 232: 379–392. http://dx.doi.org/10.1016/j.epsl.2005.01.030.

Terasaki H, Frost DJ, Rubie DC, and Langenhorst F (2007) The interconnectivity of Fe–O–S liquid in polycrystalline silicate perovskite at lower mantle conditions. *Physics of the Earth and Planetary Interiors* 161: 170–176.

Terasaki H, Frost DJ, Rubie DC, and Langenhorst F (2008) Percolative core formation in planetesimals. *Earth and Planetary Science Letters* 273: 132–137.

Terasaki H, Urakawa S, Funakoshi K, et al. (2009) In situ measurement of interfacial tension of Fe–S and Fe–P liquids under high pressure using X-ray radiography and tomography techniques. *Physics of the Earth and Planetary Interiors* 174: 220–226.

Terasaki H, Urakawa S, Rubie DC, et al. (2012) Interfacial tension of Fe–Si liquid at high pressure: Implications for liquid Fe-alloy droplet size in magma oceans. *Physics of the Earth and Planetary Interiors* 174: 220–226.

Thibault Y and Walter MJ (1995) The influence of pressure and temperature on the metal–silicate partition coefficients of nickel and cobalt in a model C1 chondrite and implications for metal segregation in a deep magma ocean. *Geochimica et Cosmochimica Acta* 59: 991–1002.

Thomas PC, Parker JW, McFadden LA, et al. (2005) Differentiation of the asteroid Ceres as revealed by its shape. *Nature* 437: 224–226.

Thomson W and Tait PG (1883) *Treatise on Natural Philosophy*. Cambridge: Cambridge University Press, p. 527.

Tonks WB and Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100: 326–346.

Tonks WB and Melosh HJ (1993) Magma ocean formation due to giant impacts. *Journal of Geophysical Research* 98: 5319–5333.

Touboul M, Kleine T, Bourdon B, Palme H, and Wieler R (2007) Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* 450: 1206–1209.

Treiman AH, Jones JH, and Drake MJ (1987) Core formation in the Shergottite parent body and comparison with the Earth. *Journal of Geophysical Research* 92: E627–E632.

Tschauner O, Zerr A, Specht S, Rocholl A, Boehler R, and Palme H (1999) Partitioning of nickel and cobalt between silicate perovskite and metal at pressures up to 80 GPa. *Nature* 398: 604–607. http://dx.doi.org/10.1038/19287.

Tsuno K, Frost DJ, and Rubie DC (2013) Simultaneous partitioning of silicon and oxygen into the Earth's core during early Earth differentiation. *Geophysical Research Letters* 40: 66–71.

Tucker JM and Mukhopadhyay S (2014) Evidence for multiple magma ocean outgassing and atmospheric loss episodes from mantle noble gases. *Earth and Planetary Science Letters* 393: 254–265.

Tuff J, Wood BJ, and Wade J (2011) The effect of Si on metal–silicate partitioning of siderophile elements and implications for the conditions of core formation. *Geochimica et Cosmochimica Acta* 75: 673–690.

Turan JW (1984) Nucleosynthesis. *Annual Review of Nuclear and Particle Science* 34: 53–97.

Turcotte DL and Schubert G (2002) *Geodynamics*. Cambridge: Cambridge University Press.

Urey HC (1952) *The Planets: Their Origin and Development*, p. 245. New Haven: Yale University Press.

von Bargen N and Waff HS (1986) Permeabilities, interfacial areas and curvatures of partially molten systems: Results of numerical computations of equilibrium microstructures. *Journal of Geophysical Research* 91: 9261–9276.

Wade J and Wood BJ (2005) Core formation and the oxidation state of the Earth. *Earth and Planetary Science Letters* 236: 78–95.

Wade J, Wood BJ, and Tuff J (2012) Metal–silicate partitioning of Mo and W at high pressures and temperatures: Evidence for late accretion of sulphur to the Earth. *Geochimica et Cosmochimica Acta* 85: 58–74.

Wagner C (1962) *Thermodynamics of Alloys*. Reading, MA: Addison-Wesley.

Walker D (2000) Core participation in mantle geochemistry: Geochemical Society Ingerson Lecture. *Geochimica et Cosmochimica Acta* 64: 2897–2911.

Walker D (2005) Core–mantle chemical issues. *Canadian Mineralogist* 43: 1553–1564.

Walker RJ (2009) Highly siderophile elements in the Earth, Moon and Mars: Update and implications for planetary accretion and differentiation. *Chemie der Erde-Geochemistry* 69: 101–125.

Walker D, Norby L, and Jones JH (1993) Superheating effects on metal–silicate partitioning of siderophile elements. *Science* 262: 1858–1861.

Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475: 206–209.

Walte NP, Becker JK, Bons PD, Rubie DC, and Frost DJ (2007) Liquid distribution and attainment of textural equilibrium in a partially-molten crystalline system with a high-dihedral-angle liquid phase. *Earth and Planetary Science Letters* 262: 517–532.

Walte NP, Rubie DC, Bons PD, and Frost DJ (2011) Deformation of a crystalline aggregate with a small percentage of high-dihedral-angle liquid: Implications for

core–mantle differentiation during planetary formation. *Earth and Planetary Science Letters* 305: 124–134.
Walter MJ, Newsom HE, Ertel W, and Holzheid A (2000) Siderophile elements in the Earth and Moon: Metal/silicate partitioning and implications for core formation. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 265–290. Tucson: University of Arizona Press.
Walter MJ and Thibault Y (1995) Partitioning of tungsten and molybdenum between metallic liquid and silicate melt. *Science* 270: 1186–1189.
Walter MJ and Tronnes RG (2004) Early Earth differentiation. *Earth and Planetary Science Letters* 225: 253–269.
Wänke H (1981) Constitution of terrestrial planets. *Philosophical Transactions of the Royal Society* 303: 287–302.
Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240.
Warren PH (1993) A concise compilation of petrologic information on possible pristine nonmare Moon rocks. *American Mineralogist* 78: 360–376.
Warren PH, Kallemeyn GW, and Kyte FT (1999) Origin of planetary cores: Evidence from highly siderophile elements in Martian meteorites. *Geochimica et Cosmochimica Acta* 63: 2105–2122.
Wasson JT (1985) *Meteorites: Their Record of Early Solar-System History*, p. 267. New York: W.H. Freeman.
Weber RC, Lin PY, Garnero EJ, Williams Q, and Lognonne P (2011) Seismic detection of the lunar core. *Science* 331: 309–312.
Weidenschilling SJ, Spaute D, Davis DR, Marzari F, and Ohtsuki K (1997) Accretional evolution of a planetesimal swarm. 2. The terrestrial zone. *Icarus* 128: 429–455.
Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
Wetherill GW and Stewart GR (1993) Formation of planetary embryos – Effects of fragmentation, low relative velocity, and independent variation of eccentricity and inclination. *Icarus* 106: 190–209.
Williams Q and Hemley RJ (2001) Hydrogen in the deep Earth. *Annual Review of Earth and Planetary Sciences* 29: 365418.
Wood BJ and Halliday AN (2005) Cooling of the Earth and core formation after the giant impact. *Nature* 437: 1345–1348.
Wood BJ and Halliday AN (2010) The lead isotopic age of the Earth can be explained by core formation alone. *Nature* 465: 767–770.
Wood BJ, Walter MJ, and Wade J (2006) Accretion of the Earth and segregation of its core. *Nature* 441: 825–833. http://dx.doi.org/10.1038/nature04763.
Yoder CF (1995) Astrometric and geodetic properties of Earth and the solar system. In: Ahrens TJ (ed.) *Global Earth Physics*, pp. 1–31. Washington, DC: American Geophysical Union.
Yoshino T, Price JD, Wark DA, and Watson EB (2006) Effect of faceting on pore geometry in texturally equilibrated rocks: Implications for low permeability at low porosity. *Contributions to Mineralogy and Petrology* 152: 169–186. http://dx.doi.org/10.1007/s00410-006-0099-y.
Yoshino T, Walter MJ, and Katsura T (2003) Core formation in planetesimals triggered by permeable flow. *Nature* 422: 154–157.
Young GA (1965) *The Physics of the Base Surge*. White Oak, MD: U.S. Naval Ordnance Laboratory. http://www.nature.com/nature/journal/v259/n5540/abs/259190a0.html.
Ziegler K, Young ED, Schauble EA, and Wasson JT (2010) Metal–silicate silicon isotope fractionation in enstatite meteorites and constraints on Earth's core formation. *Earth and Planetary Science Letters* 295: 487–496.

This page intentionally left blank

9.04 Magma Oceans and Primordial Mantle Differentiation

V Solomatov, Washington University in St. Louis, St. Louis, MO, USA

© 2015 Elsevier B.V. All rights reserved.

9.04.1	**Earth Accretion and the Giant Impact Hypothesis**	82
9.04.2	**Geochemical Evidence for Magma Ocean**	83
9.04.3	**Thermal Structure of a Convecting Magma Ocean**	84
9.04.3.1	Adiabats	84
9.04.3.2	Upper Mantle	85
9.04.3.3	Lower Mantle	85
9.04.4	**Viscosity of the Magma Ocean**	87
9.04.5	**Convection in the Magma Ocean**	88
9.04.5.1	Convective Heat Flux	88
9.04.5.2	Convective Velocities	90
9.04.6	**Fractional Versus Equilibrium Crystallization**	90
9.04.6.1	How Are Crystals Suspended by Convection?	90
9.04.6.2	Energetics of Convective Suspensions	91
9.04.6.3	Conditions for Fractional Crystallization	91
9.04.6.4	Conditions for Equilibrium Crystallization	92
9.04.7	**Crystal Size in the Magma Ocean**	92
9.04.7.1	What Processes Control the Crystal Size in the Magma Ocean?	92
9.04.7.2	Nucleation	92
9.04.7.3	Ostwald Ripening	94
9.04.8	**Crystallization Beyond the Rheological Transition**	94
9.04.9	**The Last Stages of Crystallization**	95
9.04.9.1	Cessation of Suspension	95
9.04.9.2	Formation of Thin Crust Within the Thermal Boundary Layer	96
9.04.9.3	Cessation of Liquid-State Convection	96
9.04.9.4	Percolation	96
9.04.9.5	Crystal–Melt Density Inversions	97
9.04.9.6	Remelting due to Melt Extraction	97
9.04.9.7	Solid-State Convection	97
9.04.10	**Magma Oceans on Other Planets**	98
9.04.11	**Summary**	98
References		99

Nomenclature

A	Surface area of magma ocean
a	Prefactor in nucleation function
a_u	Prefactor in scaling law for velocity
B	Exponential coefficient in nucleation function
c_p	Isobaric specific heat
c'_p	Apparent isobaric specific heat in two-phase regions
D	Diffusion coefficient
d	Crystal diameter
d_crit	Crystal diameter separating equilibrium and fractional crystallization
d_e	Critical crystal diameter for equilibrium crystallization
d_f	Critical crystal diameter for fractional crystallization
d_nucl	Crystal diameter after nucleation
d_ost	Crystal diameter controlled by Ostwald ripening
d'_ost	Crystal diameter at late stages of magma ocean crystallization
E	Activation energy of solid
E_l	Activation energy of melt
f_ϕ	Hindering settling function
F	Heat flux
g	Acceleration due to gravity
G	Gibbs energy
k_B	Boltzmann's constant
l	Mixing length
L	Depth of magma ocean
L'	Depth of shallow magma ocean
M	Mass of magma ocean
n_i	Mole fraction of component i
n_i^l	Mole fraction of component i in liquid
n_i^s	Mole fraction of component i in solid
n_mw	Mole fraction of magnesiowüstite
P	Pressure
Pr	Prandtl number
R	Gas constant

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00155-X

Ra	Rayleigh number
Ra_*	Critical Rayleigh number for transition to no-rotation regime
T	Temperature
$\dot{T}$	Cooling rate
t	Time
T'	Supercooling
Ta	Taylor number
T_{ad}	Adiabatic temperature
t_{conv}	Crystallization time due to solid-state convection
T_e	Effective temperature
T_{liq}	Liquidus temperature
T_m	Magma ocean temperature
t_{ost}	Ostwald ripening time in deep magma ocean
t'_{ost}	Ostwald ripening time in shallow magma ocean
T_{per}	Melting temperature of perovskite
t_{RT}	Rayleigh–Taylor instability time
T_s	Surface temperature
t_s	Settling time
T_{sol}	Solidus temperature
u_0	Amplitude of convective velocity
u_D	Darcy velocity
u_{perc}	Percolation velocity of melt
u_s	Settling velocity of crystals
V	Volume of magma ocean
W	Mechanical work per unit time
x^*	Velocity coefficient for turbulent boundary layer
z	Depth
α	Thermal expansion
α'	Apparent thermal expansion in two-phase regions
α_η	Coefficient in melt fraction-dependent viscosity
δ	Thermal boundary layer thickness
ΔH	Enthalpy change upon melting
ΔS	Entropy change upon melting
ΔS_{per}	Entropy change upon melting of perovskite
ΔT_{rh}	Rheological temperature scale
$\Delta\rho$	Crystal–melt density difference
$\Delta\rho_{RT}$	Density difference for Rayleigh–Taylor instability
ϵ	Efficiency factor
ζ	Ratio of supersaturation to supercooling
η	Viscosity
η_l	Viscosity of melt
η_s	Viscosity of solid
θ	Parameter in analytical solution for nucleation
κ	Coefficient of thermal diffusivity
λ	Aspect ratio of mean flow
μ_i^l	Chemical potential of component i in liquid
μ_i^s	Chemical potential of component i in solid
$\mu_i^l(0)$	Chemical potential of pure component i in liquid state
$\mu_i^s(0)$	Chemical potential of pure component i in solid state
ν	Kinematic viscosity
ρ	Density
σ	Surface energy
σ_{app}	Apparent surface energy
σ_{SB}	Stefan–Boltzmann constant
τ_T	Convective stress scale
τ^*	Skin friction
ϕ	Crystal fraction
ϕ_l	Melt fraction
ϕ_m	Maximum packing fraction
Φ	Energy release per unit time
Ω	Angular velocity

9.04.1 Earth Accretion and the Giant Impact Hypothesis

The idea that the early Earth could have been substantially molten is not new. Lord Kelvin's famous estimate of the age of the Earth was based on the assumption that the Earth was once completely molten and that the present-day surface heat flux resulted from the cooling of the molten Earth (Thomson, 1864). Later, isotopic dating showed that his estimates of the Earth's age were far too short. Radiogenic heating and mantle convection proved to play a major role in the thermal history of the Earth. The assumption of initially molten Earth did not find much observational or theoretical support and was not accepted either. The theory of Earth formation, which dominated in the middle of the twentieth century, was quite the opposite to Lord Kelvin's theory and can be characterized as soft accretion: Earth accreted from a relatively uniform influx of small particles, which gently settled on Earth without causing any 'harm.' Melting, necessary for core formation, was thought to be caused by radiogenic heating later in the Earth's history (Elsasser, 1963).

The first observational evidence that accretion was not so soft came from the Moon. The analysis of samples returned by the Apollo missions suggested a large-scale differentiation of the Moon, which was most easily explained by the crystallization of a lunar magma ocean (Longhi, 1980; Ringwood and Kesson, 1976; Warren, 1985; Wood, 1975; Wood et al., 1970). This led Hostetler and Drake to propose as early as in 1980 that, perhaps, the Earth and other planets were molten in their early histories as well (Hostetler and Drake, 1980).

That Earth accretion was very energetic became obvious upon the realization that the Earth-forming planetesimals also grew with time, like the Earth itself. It was shown that at the latest stages of planetary accretion, the population of Earth-forming planetesimals was very likely to include large bodies. Initially, the largest size was estimated at 10^{-3} of the Earth's mass or about 1000 km in diameter, based on coagulation theory and the observed planetary obliquities, which were assumed to be affected by large impacts (Safronov, 1964, 1978). Numerical investigations showed that the largest impactors could have reached about 0.1 of the Earth's mass, which is equivalent to the size of Mars (Safronov, 1964, 1978; Weidenschilling et al., 1997; Wetherill, 1975, 1985, 1990). Mars-sized impactors had enough energy to melt and partially vaporize the Earth (Benz and Cameron, 1990; Kaula, 1979; Melosh, 1990; Safronov, 1964, 1978).

The giant impact hypothesis suggested a successful explanation for the origin of the Moon, its composition, and the

angular momentum of the Earth–Moon system (Benz et al., 1986, 1987, 1989; Cameron, 1997; Canup, 2004, 2008; Canup and Agnor, 2000; Canup and Asphaug, 2001; Canup and Esposito, 1996; Ida et al., 1997; Newsom and Taylor, 1989; Stevenson, 1987). The conclusion that seems inevitable is that the Earth was born as result of rather 'violent accretion' during which it was substantially melted. An interesting description of this paradigm shift can be found in a short review by Drake (2000).

Although the kinetic energy of giant impacts was undoubtedly the largest energy source, there are other factors that contributed to Earth heating and melting. Numerical models showed that thermal blanketing effects of the atmosphere on the growing Earth helped to retain the heat during the accretion and could have maintained the surface temperature above the solidus (Abe and Matsui, 1986; Hayashi et al., 1979; Matsui and Abe, 1986; Nakazawa et al., 1985; Zahnle et al., 1988). Core formation released energy corresponding to about 2000 K heating of the entire Earth (Flasar and Birch, 1973). Some energy sources were active on a timescale of about $\sim 10^6$ years, which is much shorter than the timescale of planetary formation $\sim 10^7$–10^8 years (Canup and Agnor, 2000; Chambers and Wetherill, 1998; Wetherill, 1990). These include radiogenic heating by short-lived isotopes such as ^{26}Al and ^{60}Fe (Mostefaoui et al., 2005; Shukolyukov and Lugmair, 1993; Srinivasan et al., 1999; Urey, 1955; Yoshino et al., 2003) and electromagnetic induction heating (Sonett et al., 1968). These energy sources contributed to heating and differentiation in both the proto-Earth and the Earth-forming planetesimals during the early stages of planetary formation.

Although the accretion was a stochastic process and the Earth was melted to some degree many times by bodies of varying sizes, the major event was probably the Moon-forming impact at the end of accretion as suggested by numerical simulations (Canup, 2004; Canup and Asphaug, 2001). Such an impact seems to explain best the properties of the Earth–Moon system including its angular momentum, the masses of the Earth and Moon, and the iron depletion of the Moon. After the impact, the mantle temperature varies significantly – from 2000 to 10^4 K, with some parts of the mantle being completely molten and others remaining solid. These estimates are consistent with earlier simulations (Melosh, 1990; Pierazzo et al., 1997; Tonks and Melosh, 1993). Unfortunately, the numerical simulations can look only at the first hours of evolution (1 day at most). It is hard to predict the 'final' temperature distribution – after all major compositional and thermal heterogeneities settle down. This must be a very fast process because the stresses associated with these heterogeneities are huge and might even exceed the ultimate strength of solid rocks, which is of the order of 1–2 GPa (Davies, 1990). The gravitational energy release due to the redistribution of thermal heterogeneities in the mantle and segregation of the remaining iron increases the temperature further by probably another several hundred degrees (Tonks and Melosh, 1990, 1992). Thus, at the end of the large-scale redistribution of density heterogeneities, the hottest mantle material is completely molten, and a small fraction of the coldest material remains solid and accumulate at the bottom of the magma ocean (the solidus temperature at the base of the mantle is around 4000–5000 K; Holland and Ahrens, 1997; Zerr et al., 1998).

Interestingly, Canup's (2004) simulations show that the material that is heated most ($>10^4$ K) is the iron core of the impactor. It quickly sinks into the core of the proto-Earth, thus suggesting that the Earth's core was likely to be very hot. A substantial fraction of the energy of the superheated core is expected to be quickly transferred to the mantle (Ke and Solomatov, 2006, 2009). This may cause an additional melting of the Earth's mantle.

Another important and yet poorly quantified energy source is tidal heating. Given the fact that the Moon was very close to the Earth right after the Moon formation, it is reasonable to expect that tidal heating contributed significantly to sustaining the magma ocean after the Moon-forming event (Sears, 1993; Zahnle et al., 2007).

The term 'magma ocean' has been widely used to describe the initial molten state of the Earth. Originally, this term was applied to the lunar magma ocean and was defined as "a global, near-surface shell of magma, tens or hundreds of kilometers thick" (Warren, 1985). Warren warned about the shortcomings of this term (he proposed a different term, 'magmasphere,' but it is not used as widely as 'magma ocean'): "Among other things, "ocean" implies the system is virtually 100% liquid, with a mainly gas/liquid upper surface and a water-like viscosity of the order of 10^{-2} poise." It has been realized that the structure of the early molten mantle of the Earth and other planets can be quite complex: The molten Earth may have no stable crust in the early times but may develop a thin crust later on while still remaining molten over hundreds of kilometers below the crust; the molten layer may form not only near the surface but also in the middle of the mantle and even at the base of the mantle; and the degree of melting can vary both radially and laterally. Here, we will use the term 'magma ocean' in a broad sense – to describe a partially or a completely molten layer within the Earth's mantle whose motions are controlled largely by liquid viscosity. This means that it can be underground and it can have a relatively large viscosity.

9.04.2 Geochemical Evidence for Magma Ocean

An early global magma ocean would undoubtedly have left some record of its existence in geochemical data. Indeed, early studies showed that segregation of crystals in a deep magma ocean seemed to produce the major element composition of the upper mantle of the Earth (defined by the spinel-perovskite phase transition around 23 GPa) and in particular could explain the elevated Mg/Si ratios in the upper mantle (Agee and Walker, 1988; Herzberg and Gasparik, 1991; Ohtani, 1985; Ohtani and Sawamoto, 1987). However, geochemical models of differentiation of a terrestrial magma ocean showed that any substantial segregation of perovskite crystals in the lower mantle would drive the ratios of minor and trace elements well outside of their observed range (Kato et al., 1988a,b; McFarlane and Drake, 1990; McFarlane et al., 1994; Ringwood, 1990). On the other hand, it was pointed out that the partition coefficients were not well constrained at realistic temperatures, pressures, and compositions of magma oceans and it was premature to draw any conclusions (Presnall et al., 1998). Later studies showed that fractionation of small amounts of Ca-perovskite in addition to Mg-perovskite can increase the allowed amount

of fractionation of perovskite. Possible amounts of perovskite fractionation are in the range from 5% to 15% (Corgne and Wood, 2002; Corgne et al., 2005; Hirose et al., 2004; Liebske et al., 2005a; Walter et al., 2004; Walter and Trønnes, 2004). Ito et al. (2004) suggested that it can be as high as 40%. The third phase, magnesiowüstite (or ferropericlase), does not accommodate trace elements well so that its fractionation is not expected to have any significant effect on elemental ratios (Walter et al., 2004).

The analysis of isotopic data for ^{176}Hf–^{176}Lu and ^{147}Sm–^{143}Nd systems indicates that continental crust formed very early, perhaps in the first 200 million years of evolution, which might require a magma ocean (Albarède et al., 2000; Amelin et al., 1999; Bennett et al., 1993; Bizzarro et al., 2003; Bowring and Housh, 1995; Collerson et al., 1991; Harrison et al., 2005). Isotopic data on compositions of ^{142}Nd produced by short-lived ^{146}Sm (Boyet and Carlson, 2005; Caro et al., 2003) indicate even shorter timescales, within 30 million years, for mantle differentiation.

A simultaneous modeling of several isotopic systems can potentially provide a very tight constraint on mantle differentiation (e.g., Halliday, 2004). Caro et al. (2005) argued that superchondritic Sm/Nd ratios of early Archean rocks can be reconciled with a nearly chondritic Lu/Hf ratio observed in zircons from the same place and of the same age by extraction of about 0.3 wt% of melt from the upper mantle. Differentiation at small degrees of melting was also proposed earlier to explain the ratio of major and minor elements (Gasparik and Drake, 1995).

Another argument for an early molten state of the Earth comes from the fact that core formation on Earth as well as other planetary bodies happened fast, perhaps within 30–50 Myr of the solar system history, which is inferred from the analysis of the short-lived ^{182}Hf–^{182}W system (Foley et al., 2005; Halliday, 2004; Halliday et al., 1996; Jacobsen, 2005; Kleine et al., 2002; Lee and Halliday, 1995; Yin et al., 2002). This requires some degree of melting of the Earth's mantle very early in planetary history (Stevenson, 1990). To what extent the Earth was molten during core formation is not constrained by Hf–W data. The depth of the molten layer of the Earth can be inferred from the analysis of siderophile elements. Laboratory experiments on partition coefficients of siderophile elements suggest that metal/silicate chemical equilibrium was established around 28 GPa and 2200 K, which can be interpreted as the bottom of a magma ocean (Li and Agee, 1996; Righter, 2003, 2011; Righter and Drake, 1997; Righter et al., 1997; Siebert et al., 2012). According to this model, iron delivered by impacts first accumulates at the bottom of a magma ocean and then sinks through the mostly solid layers below the magma ocean (Elsasser, 1963; Karato and Murthy, 1997a,b; Solomatov, 2000; Stevenson, 1981, 1990). The pressure at the bottom of the magma ocean might be noticeably different from the apparent equilibration pressure because metal/silicate equilibrium was established during a 'rainfall' of iron droplets rather than in the iron layer at the base of magma ocean (Rubie et al., 2003). Besides, the bottom of the magma ocean may not be well described as a simple boundary but rather as an extensive partially molten region with variable melt fraction and a complicated structure and dynamics.

The geochemical constraints can be summarized as follows. Only a small amount of perovskite could have been fractionated via crystal–melt segregation (~10%). Chemical differentiation was likely to occur in the upper parts of the mantle via percolation of melt in a partially molten mantle at small melt fractions. The timescales for core formation and mantle differentiation appear to be very close to each other, and both are close to the accretion timescale constrained by numerical simulations (Canup and Agnor, 2000; Chambers and Wetherill, 1998; Wetherill, 1990). This suggests that core formation, mantle differentiation, and accretion might have been occurring simultaneously. The discussion in the succeeding text is based largely on Solomatov and Stevenson (1993a,b,c), Solomatov et al. (1993), and Solomatov (2000).

9.04.3 Thermal Structure of a Convecting Magma Ocean

9.04.3.1 Adiabats

In vigorously convecting systems, such as magma oceans, the temperature distribution is nearly adiabatic and isentropic. In one-phase systems, such as a completely molten or a completely solid layer, the equation for an adiabat is

$$\frac{\mathrm{d}T}{\mathrm{d}P}=\frac{\alpha T}{\rho c_{\mathrm{p}}} \qquad [1]$$

where T is the temperature, P is the pressure, α is the coefficient of thermal expansion, c_{p} is the isobaric specific heat, and ρ is the density.

In two-phase systems, the effects of phase changes need to be considered (Abe, 1993, 1995, 1997; Miller et al., 1991a,b; Solomatov and Stevenson, 1993b). In the absence of chemical differentiation and assuming thermodynamic equilibrium, the volume fractions as well as compositions of each mineral phase depend only on T and p. Thus, the system can be described with the help of two thermodynamic variables, T and p. All thermodynamic parameters such as the coefficient of thermal expansion and the isobaric specific heat can be obtained the same way as for a one-phase system. In terms of two-phase parameters α' and c_{p}', the equation for the adiabat in the temperature range between liquidus and solidus can be written as follows:

$$\frac{\mathrm{d}T}{\mathrm{d}P}=\frac{\alpha' T}{\rho c_{\mathrm{p}}'} \qquad [2]$$

The order-of-magnitude values of α' and c_{p}' are

$$\alpha' \sim \alpha + \frac{\Delta\rho}{\rho\left(T_{\mathrm{liq}}-T_{\mathrm{sol}}\right)} \qquad [3]$$

and

$$c_{\mathrm{p}}' \sim c_{\mathrm{p}} + \frac{\Delta H}{T_{\mathrm{liq}}-T_{\mathrm{sol}}} \qquad [4]$$

where ΔH is the specific enthalpy change upon melting, $\Delta\rho$ is the density difference between solid and liquid, T_{liq} is the liquidus temperature, and T_{sol} the solidus temperature. Usually, the second term dominates, $\alpha' \gg \alpha$, and $c_{\mathrm{p}}' \gg c_{\mathrm{p}}$. Also, α' is affected more strongly than c_{p}' so that the two-phase adiabat is usually steeper than either one-phase adiabat (see Solomatov and Stevenson, 1993b, for the examples when this is not the case). Note that the situation is different for the atmosphere

where the so-called 'wet' adiabat (due to condensation of water) is always less steep than the 'dry' adiabat (e.g., Goody, 1995). Also, because $\alpha' \gg \alpha$ and $c_p' \gg c_p$, the adiabatic gradient, eqn [2], is approximately equal to the Clapeyron slope of the dominant crystallizing phase and the adiabats tend to align with the curves of constant crystal fraction (e.g., solidus and liquidus).

There are several approaches to calculate the adiabats in the magma ocean. The simplest approach is to parameterize the melt fraction as a function of temperature and pressure without considering what phases crystallize and what composition they are (Abe, 1997; Miller et al., 1991a,b). The most rigorous approach would involve building a self-consistent thermodynamic model, which would use the available database for both solid and liquid silicates (e.g., Asimow et al., 1997; Ghiorso, 1997). An intermediate approach is to choose some simple system consisting of only a few components and treat it self-consistently (Solomatov and Stevenson, 1993b).

9.04.3.2 Upper Mantle

Solomatov and Stevenson (1993b) considered two types of three-component systems to describe thermodynamics of the upper mantle. Here, we only consider the eutectic-like system, which seems to be most accurate at low pressures ($P < 10$ GPa). In this system, the three components, olivine, orthopyroxene, and clinopyroxene, are completely insoluble in each other in the solid state. The Gibbs free energy of the system (solid plus melt) is written as

$$G = \sum_{i=1}^{3} \left(n_i^{\mathrm{s}} \mu_i^{\mathrm{s}} + n_i^{\mathrm{l}} \mu_i^{\mathrm{l}} \right) \quad [5]$$

where μ_i^{s} and μ_i^{l} are the chemical potentials of the component i ($i = 1, 2, 3$) in the solid and liquid, respectively, n_i^{s} and n_i^{l} are the mole fractions of the solid and liquid component i, respectively,

$$n_i^{\mathrm{s}} + n_i^{\mathrm{l}} = n_i \quad [6]$$

and n_i is a total mole fraction of the component i in the system:

$$\sum_{i=1}^{3} n_i = 1 \quad [7]$$

The solid phases are assumed to be insoluble in each other. Thus, the chemical potentials of the solid phases are the same as the chemical potentials $\mu_i^{\mathrm{s}(0)}$ of the pure components:

$$\mu_i^{\mathrm{s}} = \mu_i^{\mathrm{s}(0)} \quad [8]$$

The liquid phases are assumed to form ideal solutions:

$$\mu_i^{\mathrm{l}} = \mu_i^{\mathrm{l}(0)} + RT \ln N_i^{\mathrm{l}} \quad [9]$$

where $\mu_i^{\mathrm{l}(0)}$ is the chemical potential of a pure component i in liquid state and N_i^{l} is the mole fraction of the component i in the liquid:

$$N_i^{\mathrm{l}} = \frac{n_i^{\mathrm{l}}}{\sum_{i=1}^{3} n_i^{\mathrm{l}}} \quad [10]$$

The requirement of thermodynamic equilibrium,

$$\mu_i^{\mathrm{l}} = \mu_i^{\mathrm{s}} \quad [11]$$

allows us to uniquely determine the composition of liquid and solid (n_i^{l} and n_i^{s}) and the melt fraction (the mole melt fraction is simply $\sum_{i=1}^{3} n_i^{\mathrm{l}}$) at any given pressure and temperature. All thermodynamic parameters of the combined crystal–melt system, such as the specific heat and the coefficient of thermal expansion, can be calculated from the Gibbs energy using standard thermodynamic relationships.

Figure 1 shows adiabats for this system at $P < 10$ GPa. Variation of melt fraction between liquidus and solidus is given in Figure 2 along with the experimental data from McKenzie and Bickle (1988). The melting temperatures, their gradients, and the fractions of the three components are slightly adjusted to fit the experimental data. Figure 2 shows that the data follow the expected general pattern of an ideal three-component eutectic system. Near the solidus, all three solid phases undergo eutectic melting characterized by a jump in the melt fraction, and one component (clinopyroxene) is completely liquid at the end of eutectic melting. At some melt fraction (around 30%), the slope changes because the second component is completely molten (orthopyroxene). Olivine remains the only solid phase all the way to liquidus.

9.04.3.3 Lower Mantle

The three components in the lower mantle are periclase, MgO; wüstite, FeO; and perovskite, $MgSiO_3$. Perovskite has only a small amount of iron and is considered to be a pure $MgSiO_3$. The system is very similar to the three-component system described earlier for the upper mantle. The main difference is that components 1 and 2 (periclase and wüstite) are assumed to form an ideal solid solution with each other (magnesiowüstite).

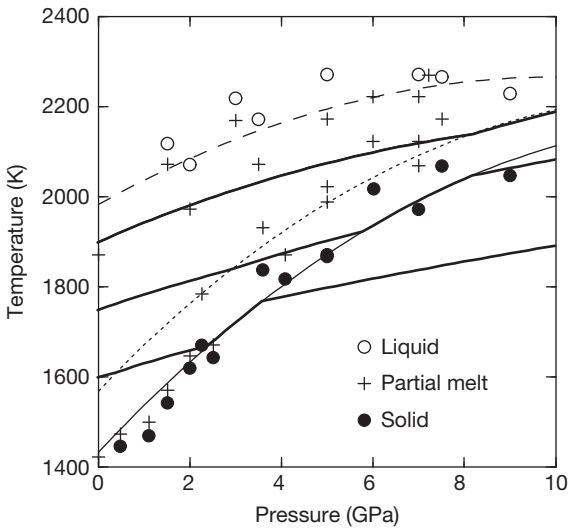


Figure 1 Adiabats in the upper mantle (thick solid lines). Liquidus (dashed line), solidus (solid line), and the beginning of crystallization of orthopyroxene (dotted line) are shown together with experimental data for peridotites (Ito and Takahashi, 1987; McKenzie and Bickle, 1988; Scarfe and Takahashi, 1986). Reproduced from Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tuscon, AZ: University of Arizona Press.

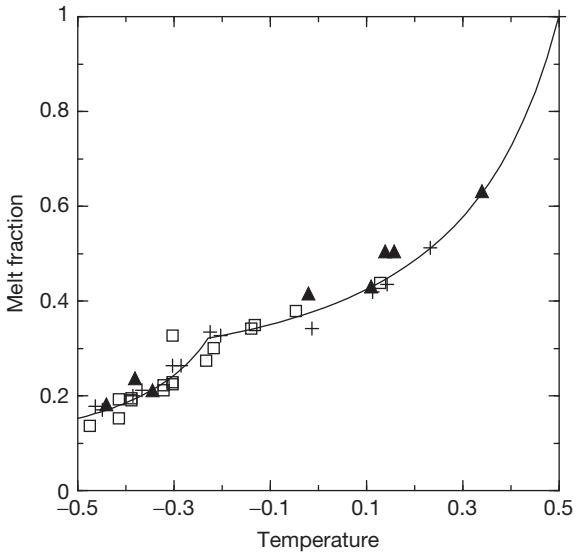


Figure 2 Variation of melt fraction between liquidus ($T=-0.5$) and solidus ($T=0.5$). The experimental data are from McKenzie and Bickle (1988): $0 \le P \le 0.5$ GPa (crosses), $0.5 < P \le 1.5$ GPa (squares), $P > 1.5$ GPa (triangles). The theoretical curve is calculated for an idealized eutectic system consisting of three immiscible components: olivine, orthopyroxene, and clinopyroxene. It shows a typical behavior of this type of systems: a steplike melting at the solidus during which one component – clinopyroxene – melts completely and an abrupt slope change around ~30% melt fraction where the second component – orthopyroxene – melts completely so that only olivine crystals are present until the temperature reaches liquidus. Reproduced from Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tuscon, AZ: University of Arizona Press.

Thus, the chemical potentials of the solid phases are described as follows:

$$\mu_i^s = \mu_i^{s(0)} + RT \ln N_i^s, \quad i = 1, 2 \qquad [12]$$

$$\mu_3^s = \mu_3^{s(0)} \qquad [13]$$

where N_i^s is the mole fraction of the component i in the solid:

$$N_i^s = \frac{n_i^s}{\sum_{i=1}^{3} n_i^s} \qquad [14]$$

Figure 3 shows an example of calculations for the lower mantle. The melting temperatures of $MgSiO_3$, MgO, and FeO are based on laboratory experiments and extrapolation to high pressures by Boehler (1992) and Zerr and Boehler (1993, 1994).

The liquidus and solidus curves predicted by this model for the lower mantle are steeper than single phase adiabats. The solidus is very close to the one estimated by Holland and Ahrens (1997) and Zerr et al. (1998). The liquidus T_{liq} approximately follows the melting temperature T_{prv} of pure perovskite $MgSiO_3$ lowered by the presence of MgO and FeO:

$$T_{prv} - T_{liq} \approx \frac{k_B}{\Delta S_{prv}} n_{mw} T_{prv} \qquad [15]$$

where ΔS_{prv} is the entropy change per atom upon melting of pure perovskite and n_{mw} is the mole fraction of magnesiowüstite.

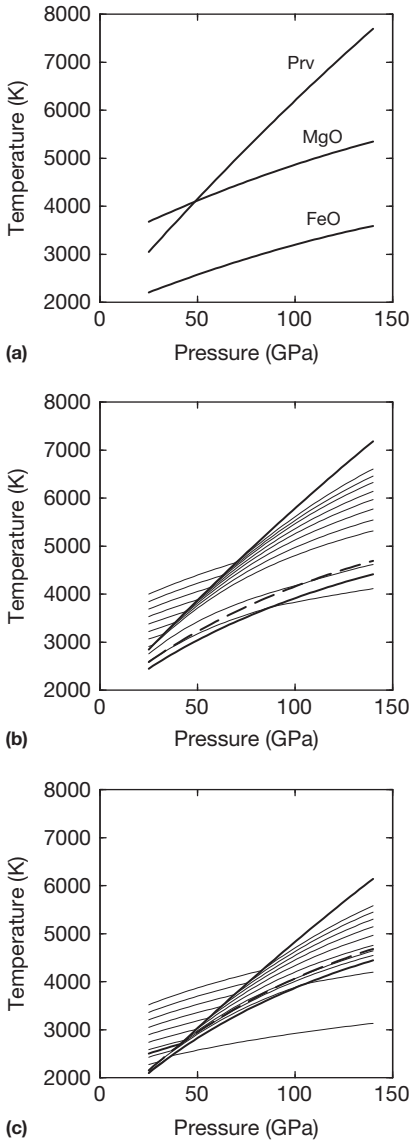


Figure 3 (a) Melting curves of perovskite, MgO and FeO (Boehler, 1992; Zerr and Boehler, 1993, 1994). (b) Adiabats in the convective magma ocean. Liquidus and solidus are shown with heavy solid line. The beginning of magnesiowüstite crystallization is shown with a dashed line. (c) An example of calculations where magnesiowüstite is the first liquidus phase at the top of the lower mantle. Reproduced from Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tuscon, AZ: University of Arizona Press.

For $n_{mw} \approx 0.3$, $\Delta S_{prv} \approx 5k_B$ (assuming that the entropy change is approximately k_B per atom, where k_B is Boltzmann's constant; Stishov, 1988) and $T_{prv} \approx 7500$ K (at the base of the mantle), we obtain $T_{prv} - T_{liq} \approx 450$ K. This shows that the liquidus strongly depends on the assumed melting temperature of perovskite. For comparison, Miller et al. (1991a,b) and Solomatov and Stevenson (1993b) assumed a lower melting temperature of perovskite based on the data by Knittle and Jeanloz (1989). The liquidus also depends on the details of the phase diagram. Abe's (1997) lower liquidus is related to his assumption that $MgSiO_3$

and MgO form an ideal solid solution, while in our calculations, perovskite forms eutectic-like subsystems with periclase and wüstite. Recent laboratory experiments and theoretical models significantly refined the thermodynamics of the lower mantle. In particular, they suggest a rather low solidus and liquidus temperatures in the lower mantle and with a relatively small (a few hundred degrees) gap between them at the core–mantle boundary (Andrault et al., 2012; de Koker et al., 2013; de Koker and Stixrude, 2009).

Agee (1990) and Zhang and Herzberg (1994) found that magnesiowüstite is the liquidus phase in both chondrites and peridotites just slightly below the upper mantle/lower mantle boundary. A small fraction of iron in the solid perovskite and the presence of other components (Ca-perovskite) in melt would decrease the liquidus temperature. The estimates of the entropy change are also not very accurate. As a result, magnesiowüstite can be the first liquidus phase at low pressures although at higher pressures, it would still be substituted by perovskite (**Figure 3**).

It is interesting to note that, within the uncertainties of the thermodynamics of the system, crystallization in the lower mantle may not necessarily proceed monotonically from top to bottom. In addition to the crystallizing layer at the surface, an additional layer may form somewhere in the middle of the lower mantle propagating both upward and downward (Mosenfelder et al., 2009; Solomatov and Stevenson, 1993b; Thomas et al., 2012). In combination with crystal–melt density inversion, this may lead to the formation of more than one long-lived partially molten layers at the end of magma ocean crystallization (see also **Section 9.04.9.5**).

9.04.4 Viscosity of the Magma Ocean

Viscosity is an important parameter that controls virtually all dynamic processes in the magma ocean and affects the outcome of crystallization of the magma ocean. The viscosity of the magma ocean varies in a very broad range.

At low degrees of melting, the viscosity of peridotitic magmas can be quite high, on the order of 100 Pa s (Kushiro, 1980, 1986). This is related to relatively low temperature as well as relatively felsic composition – the large concentration of silicon in the initial melts tends to increase the viscosity. The viscosity decreases with temperature very rapidly. Experimental and theoretical studies suggest that the viscosity of many near-liquidus ultramafic silicates at low pressures is around $\eta_l \sim 0.1$ Pa s (Bottinga et al., 1995; Bottinga and Weill, 1972; Dingwell et al., 2004; Liebske et al., 2005b; Persikov et al., 1997; Shaw, 1972).

There are no experimental data for the viscosity of the lower mantle. The viscosity behaves nonmonotonically with pressure because of the structural changes in the polymerized melts (e.g., Kushiro, 1980; Poe et al., 1997). The viscosity of a completely depolymerized melt weakly increases with pressure (Andrade, 1952; Gans, 1972). An increase of less than one order of magnitude can be expected along the liquidus throughout the lower mantle. Thus, the viscosity of magma oceans near the liquidus is probably around 10^{-1} Pa s. In the region between solidus and liquidus, the viscosity of pure melt (without crystals) is somewhat higher because of the lower temperatures. However, the temperature effect is small for low-viscosity liquids (<1 Pa s), which exhibit a power law rather than Arrhenius behavior (Bottinga et al., 1995; Liebske et al., 2005b). First-principles molecular dynamics simulations give similar results (Adjaoud et al., 2012; Karki and Stixrude, 2010). The value of 10^{-1} Pa s with the uncertainty of a factor of 10 will be assumed for the upper mantle as well as for the lower mantle (**Table 1**).

Many authors pointed out that magma oceans can contain substantial amounts of water (Abe et al., 2000; Ahrens, 1992; Righter et al., 1997). Although water reduces the viscosity of magmas, in the limit of high temperatures, the viscosities and diffusivities of many liquids including water are very similar (Persikov et al., 1997). Therefore, the viscosity of completely depolymerized, high-temperature hydrous magma is unlikely to be much different from that of anhydrous magma.

The most important effect on magma viscosity is the presence of crystals. The viscosity of a crystal–melt mixture changes relatively little with the solid fraction. At small volume fraction ϕ of solids, the viscosity η of the mixture can approximately be described with the help of Einstein's linear formula for diluted suspensions of perfectly spherical particles (Einstein, 1906):

$$\eta = \eta_l(1 + 2.5\phi) \quad [16]$$

Various formulas have been proposed to describe the viscosity of suspensions at large crystal fractions (Brinkman, 1952; Campbell and Forgacs, 1990; Costa, 2005; Frankel and Acrivos, 1967; Krieger and Dougherty, 1959; McBirney and Murase, 1984; Mooney, 1951; Murray, 1965; Roscoe, 1952). For example, one of the most popular expressions is (Roscoe, 1952)

$$\eta = \frac{\eta_l}{(1 - \phi/\phi_m)^{2.5}} \quad [17]$$

where ϕ_m is the maximum packing crystal fraction.

Although the previous equation does not describe the viscosity accurately near $\phi = \phi_m$, it does show an important physical aspect of solid/liquid mixtures, that is, the viscosity becomes

Table 1 Typical parameters of a deep magma ocean

Density, ρ	4×10^3	kg m^{-3}
Temperature, T	4×10^3	K
Thermal expansion, α	5×10^{-5}	K^{-1}
Thermal capacity, c_p	10^3	J kg^{-1} K^{-1}
Gravity, g	10	m s^{-2}
Crystal–melt density difference, $\Delta\rho$	300	kg m^{-3}
Viscosity of melt at liquidus, η_l	0.1	Pa s
Viscosity of solids at solidus, η_s	10^{18}	Pa s
Diffusion coefficient, D	10^{-9}	m^2 s^{-1}
Enthalpy change upon melting, $\Delta H\rho$	10^6	J kg^{-1}
Surface energy, σ	0.5	J m^{-2}
Apparent surface energy, σ_{app}	0.02	J m^{-2}
Liquidus gradient, dT_{liq}/dz	2×10^{-3}	K m^{-1}
Adiabatic gradient, dT_{ad}/dz	6×10^{-4}	K m^{-1}
Heat flux, F	10^6	J m^{-2} s^{-1}
Angular velocity, Ω	10^{-4}	s^{-1}
Magma ocean depth, L	3×10^6	m
Remaining molten layer depth, L'	5×10^5	m
Convective velocity, u_0	10	m s^{-1}
Timescale for Ostwald ripening, t_{ost}	10^6	s
Effective cooling rate in plumes, $\dot{T}$	0.2	K s^{-1}
Nucleation prefactor, a	10^{20}	m^{-3} s^{-1}
Crystal size, d	10^{-3}	m

infinite when the crystal fraction reaches ϕ_m. The deformation beyond $\phi = \phi_m$ is only possible if crystals deform via solid-state creep, which is many orders of magnitude slower.

The value of ϕ_m for ideal spheres is around 0.64. For a binary mixture in which the small spheres have just the right size to fit the space in between large spheres (about 30% smaller than the large spheres), $\phi_m = 0.86$ (McGeary, 1961). In reality, the crystal distribution is far from any hypothetical one. Moreover, the shape of crystals is not spherical. Saar et al. (2001) showed that crystal network can form at crystal fraction as low as 10% or 20% depending on the elongation (the smaller number is for a rather large ratio of the longest side of the crystal to the shortest one – over a factor of 10). In the presence of shear, the elongation can affect ϕ_m in the opposite direction. Elongated crystals can align with the flow and develop a hydrodense suspension with as much as 80% crystal fraction (Nicolas and Ildefonse, 1996).

The experiments on partial melts provide the most direct constraint on magma viscosity and ϕ_m, although the melt composition is still different from that of the magma ocean. These experiments show a rapid increase of the viscosity at high crystal fractions and $\phi_m \approx 60\%$ (Arzi, 1978; Lejeune and Richet, 1995; van der Molen and Paterson, 1979). The value of ϕ_m in the experiments is determined from a theoretical fit using equations similar to eqn [17].

Complications arise due to crystal–crystal interactions in the vicinity of ϕ_m. They are most likely to be responsible for the non-Newtonian behavior of partial melts characterized by substantial yield strength (reported values range from 10^2 to 10^6 Pa) and nonlinear relationship between stress and strain rate (Lejeune and Richet, 1995; McBirney and Murase, 1984; Pinkerton and Sparks, 1978; Robson, 1967; Ryerson et al., 1988; Shaw, 1969; Shaw et al., 1968).

The transition from fluidlike behavior to solid-like behavior occurs in a rather narrow range of crystal fraction. Thus, it can be called the 'rheological transition.' At ϕ only slightly smaller than ϕ_m, the viscosity is not much different from that of a pure melt except for corrections described by eqns [16] and [17]. At $\phi > \phi_m$, the deformation rate is controlled by the viscosity of the solid matrix reduced by the presence of melt (Cooper and Kohlstedt, 1986; Hirth and Kohlstedt, 1995a,b; Jin et al., 1994; Kohlstedt and Zimmerman, 1996; Mei et al., 2002; Mei and Kohlstedt, 2000a,b; Rutter and Neumann, 1995; van der Molen and Paterson, 1979). The dependence of the solid-state viscosity on the melt fraction is parameterized with the help of simple exponential function:

$$\eta = \eta_s \exp(-\alpha_\eta \phi) \quad [18]$$

where η_s is the viscosity of melt-free rock and α_η is a constant that depends on the creep mechanism. Mei et al. (2002) gave $\alpha_\eta = 26$ for diffusion creep and $\alpha_\eta = 31$ for dislocation creep.

9.04.5 Convection in the Magma Ocean

9.04.5.1 Convective Heat Flux

During the early stages of crystallization of the magma ocean, the viscosity is small and convection is extremely turbulent. The convective heat flux is usually calculated as

$$F_{soft} = 0.089 \frac{k(T_m - T_s)}{L} Ra^{1/3} \quad [19]$$

where

$$Ra = \frac{\alpha g (T_m - T_s) L^3}{\kappa \nu} \quad [20]$$

is the Rayleigh number, T_m is the potential temperature of the magma ocean, T_s is the surface temperature, k is the coefficient of thermal conductivity, $\kappa = k/\rho c_p$ is the coefficient of thermal diffusivity, and $\nu = \eta/\rho$ is the kinematic viscosity (Kraichnan, 1962; Siggia, 1994).

However, at very high Rayleigh numbers such as those in the magma ocean, convection changes to a regime sometimes called hard turbulence (Castaing et al., 1989; Grossmann and Lohse, 1992; Shraiman and Siggia, 1990; Siggia, 1994). By contrast, the ordinary turbulence is called soft turbulence. One of the important features of hard-turbulent convection is the existence of a large-scale circulation. This coherent motion within the highly turbulent fluid is an example of self-organization in complex systems (e.g., Nicolis and Prigogine, 1977), that is, systems that have many degrees of freedom interacting with each other. Turbulence is caused by numerous plumes originating from the upper and lower boundaries of the convective layer. The coherent circulation of the fluid emerges from random actions of these plumes. In the magma ocean, this mainly includes cold plumes originating from the surface but may also include hot plumes originating from the core–mantle boundary.

Several scaling laws have been proposed for the hard turbulence regime (Kadanoff, 2001; Siggia, 1994). Shraiman and Siggia (1990) obtained the following equation:

$$F_{hard} = 0.22 \frac{k(T_m - T_s)}{L} Ra^{2/7} Pr^{-1/7} \lambda^{-3/7} \quad [21]$$

where $Pr = \nu/\kappa$ is the Prandtl number and λ is the aspect ratio for the mean flow.

Another effect to consider is the rotation of the Earth. Although the rotation of the Earth does not affect subsolidus mantle dynamics (Coriolis forces are negligible compared to viscous forces), the dynamics of low-viscosity magma ocean may be affected by rotation. How large is the effect of rotation on the heat flux? At low Rayleigh numbers (soft turbulence), the heat flux depends on the rotation as follows:

$$F \sim \frac{k(T_m - T_s)}{L} Ra^3 Ta^{-2} \quad [22]$$

where

$$Ta = \frac{4\Omega^2 L^4}{\nu^2} \quad [23]$$

is the Taylor number (Boubnov and Golitsyn, 1986, 1990; Canuto and Dubovikov, 1998).

At high Rayleigh numbers, the effect of rotation becomes negligible and the heat flux is the same as in the absence of rotation. The critical Rayleigh number for the transition to the no-rotation regime is approximately the one at which the two scaling laws (with and without rotation) give the same heat flux. In the soft turbulence regime, the requirement that the heat flux [19] is the same as [22] gives the Rayleigh number above which rotation does not affect the heat flux (see also Canuto and Dubovikov, 1998):

$$Ra_* \sim Ta^{3/4} \quad [24]$$

For the magma ocean, $Ta \sim 10^{25}$. This gives $Ra_* \sim 10^{19}$, which is much smaller than the typical values $Ra \sim 10^{28}$–10^{29}. Although these estimates are for soft turbulence, the difference between soft turbulence and hard turbulence is relatively small. Thus, it is reasonable to assume that rotation does not have any significant effect on the heat flux in the hard turbulence regime.

Application of the hard turbulence regime to magma oceans requires substantial extrapolation: Laboratory experiments have only been performed on helium ($Pr \sim 1$) to about $Ra \sim 10^{17}$, while in the magma ocean, $Ra \sim 10^{28}$–10^{29} and $Pr \sim 10^2$–10^3. At very high Rayleigh numbers, convection was expected to enter a new regime of turbulent convection (Kraichnan, 1962; Siggia, 1994). So far, experiments did not show any evidence of such an 'ultrahard' regime (Glazier, 1999; Niemela et al., 2000). Thus, hard turbulence is probably applicable to the extreme conditions of magma oceans.

The spherical geometry of the magma ocean is different from the geometries in laboratory experiments. However, like in other convection problems, the changes in the scaling laws due to spherical geometry are expected to be minor. A major uncertainty associated with spherical geometry is probably the aspect ratio λ of the mean flow. The simplest assumption is that $\lambda \sim 1$, that is, the horizontal scale of the mean flow is of the order of the depth of the magma ocean. This parameter can be effected by rotation.

Temperature-dependent viscosity does not seem to play a significant role in controlling the heat flux. When the potential temperature is close to liquidus, the temperature contrast in the thermal boundary layer is 200–600 K depending on the scaling law. According to Liebske et al. (2005b), the corresponding viscosity contrast is at most one order of magnitude. This is too small to have a significant effect on the heat transport (Solomatov, 1995a).

Does radiative heat transport play a role in the thermal boundary layer near the surface? The thickness $\delta = k(T_m - T_s)/F$ of the thermal boundary layer is of the order of 1 cm. The free path of infrared photons in silicates is smaller, of the order of 1 mm (Clark, 1957; Shankland et al., 1979). This means that the thermal boundary layer is not thin enough to be transparent to infrared radiation. Also, the thermal conductivity of magma does not seem to be significantly affected by the radiative heat transport. The data collected by Murase and McBirney (1973) suggest that for solid lherzolite, the thermal conductivity is not very different from that at low temperatures: $k \approx 4\,\mathrm{W\,m^{-1}\,K^{-1}}$ around 1500 °C.

The heat flux depends on the surface temperature. Since the heat flux from the magma ocean is much higher than the incoming solar radiation, the surface temperature is established not by the radiative equilibrium between the incoming radiation and outgoing radiation but by the equilibrium between the heat flux transported by convection in the magma ocean to the surface and the heat flux radiated from the surface of the magma ocean. The latter depends on the type of the atmosphere covering the magma ocean (Abe and Matsui, 1986; Hayashi et al., 1979; Matsui and Abe, 1986; Nakazawa et al., 1985; Zahnle et al., 1988). In the early stages of crystallization, the surface temperature is very high, more than 2000 K. In this case, the atmosphere is a 'silicate' one (Hayashi et al., 1979; Nakazawa et al., 1985; Thompson and Stevenson, 1988). The heat flux can be written as

$$F = \sigma_{SB} T_e^4 \quad [25]$$

where T_e is the effective temperature (the temperature that corresponds to a blackbody emitting the heat flux F) and $\sigma_{SB} = 5.67 \times 10^{-8}\,\mathrm{J\,m^{-2}\,K^{-4}}$ is the Stefan–Boltzmann constant. The magnitude of T_e depends on the mass of the silicate atmosphere. It can be smaller than the surface temperature T_s by several hundred kelvins (Thompson and Stevenson, 1988). Fortunately, the dependence on the mass of the atmosphere is weak. The simplest assumption one can make without considering the heat transfer in the atmosphere is that $T_e \approx T_s$, that is, magma ocean radiates like a blackbody with temperature T_s. With this assumption, the surface heat flux can be overestimated by a factor of $(T_s/T_e)^4$. For a moderately opaque atmosphere with $T_e \approx 1500\,\mathrm{K}$ and $T_s \approx 2000\,\mathrm{K}$, $(T_s/T_e)^4 \sim 3$.

Figure 4 shows the heat flux and the surface temperature as functions of the potential temperature of the magma ocean during the initial period of crystallization. The scaling law for

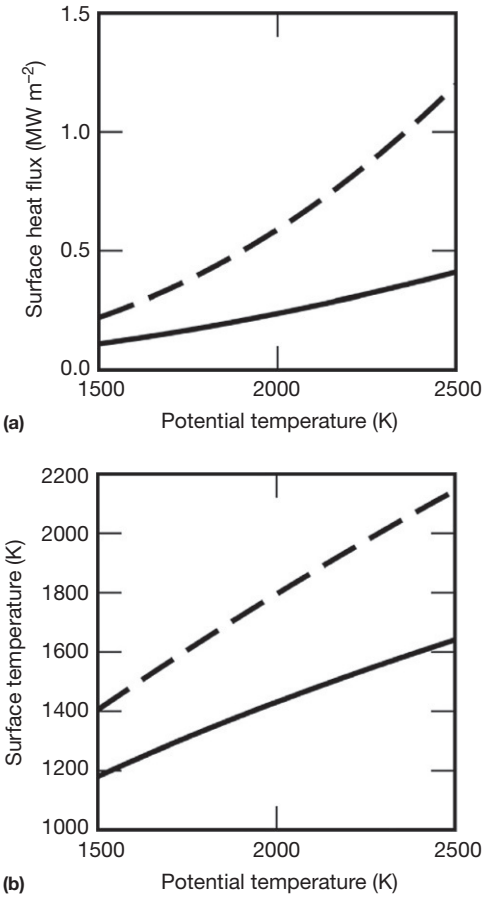


Figure 4 (a) The surface heat flux and (b) the surface temperature are shown as functions of the potential temperature calculated using scaling laws for soft turbulence convection (solid line) and hard turbulence convection (dashed line). The viscosity is $\eta_l = 0.1$ Pa s. The uncertainties associated with the viscosity are smaller than those associated with the scaling law.

hard-turbulent convection predicts that the heat flux is close to $10^6\,\mathrm{W\,m^{-2}}$. The formula for soft-turbulent convection gives lower values of the heat flux, by a factor of ~ 3. It might be worth emphasizing that these are only rough estimates. Also, a more realistic model would need to consider a coupled evolution of the atmosphere and magma ocean, which is a significantly more complex problem (Abe, 1997; Elkins-Tanton, 2011; Marcq, 2012) especially given its highly stochastic nature (Abe, 2011).

9.04.5.2 Convective Velocities

In the soft turbulence regime, the convective velocities can be estimated in terms of the mixing length l as (Kraichnan, 1962; Priestly, 1959)

$$u_0 \approx 0.6\left(\frac{\alpha g l F}{\rho c_\mathrm{p}}\right)^{1/3} \quad [26]$$

where F is the surface heat flux. The coefficient is constrained by laboratory experiments (Deardorff, 1970; Willis and Deardorff, 1974) and atmospheric measurements (Caughey and Palmer, 1979). The simplest assumption is that the mixing length is approximately equal to the depth of the magma ocean $l \sim L$ (if the magma ocean has partially crystallized beyond the rheological transition, then l should be the thickness of the remaining low-viscosity region). For $F \approx 10^6\,\mathrm{W\,m^{-2}}$ (**Figure 4**) and other parameters from **Table 1**, eqn [26] gives $u_0 \sim 4\ \mathrm{m\ s^{-1}}$ (note that the velocity is slightly larger in two-phase regions where α and c_p need to be replaced by eqns [3] and [4]).

The ratio F/ρ is almost the same for the magma ocean and the atmosphere. This explains why the previous estimate is very similar to the observed velocities in the convective boundary layer in the atmosphere (Caughey and Palmer, 1979).

The equations suggested by Shraiman and Siggia (1990) can be rewritten in the form similar to eqn [26]:

$$u_0 \approx a_u\left(\frac{\alpha g l F}{\rho c_\mathrm{p}}\right)^{1/3} \quad [27]$$

where the coefficient $a_u \approx 0.086 x^*$ and x^* is related to u_0 through the following equation:

$$x^* = 2.5 \ln\left[\frac{\rho u_0 l}{\eta}\frac{1}{x^*}\right] + 6 \quad [28]$$

The solution to the earlier transcendental equations gives $a_u \approx 5.9$, which varies weakly with $\rho u_0 l/\eta$. In the hard turbulence regime, the velocity increases by a factor of 10. This brings the estimate of the convective velocity in the magma ocean up to $40\ \mathrm{m\,s^{-1}}$. The effect of the Earth's rotation can be estimated in the same way as it was done for soft turbulence: replacing the thickness of the convective layer by a reduced length scale

$$l \sim \frac{u_0}{\Omega} \quad [29]$$

imposed by rotation explains rather well the experimentally observed reduction in the convective velocity (Boubnov and Golitsyn, 1986, 1990; Chen et al., 1989; Fernando et al., 1991; Golitsyn, 1980, 1981; Hopfinger, 1989; Hopfinger et al., 1982; Solomatov and Stevenson, 1993a).

Assuming that this length scale works for hard turbulence, Solomatov (2000) obtained

$$u_0 \approx 14\left(\frac{\alpha g F}{\rho c_\mathrm{p} \Omega}\right)^{1/2} \quad [30]$$

This gives velocities around $16\ \mathrm{m\,s^{-1}}$ with the uncertainty of a factor of 3. The value of $10\ \mathrm{m\,s^{-1}}$ can be assumed for the magma ocean (**Table 1**).

9.04.6 Fractional Versus Equilibrium Crystallization

9.04.6.1 How Are Crystals Suspended by Convection?

Early studies of convective suspensions used several approaches. Some looked at the trajectories of particles in steady flows and assumed that the particles whose trajectories are closed can be suspended indefinitely (Marsh and Maxey, 1985; Rudman, 1992; Weinstein et al., 1988). This approach cannot work for turbulent flows where closed trajectories do not exist.

Others used a phenomenological description of the balance between the downward flux of particles due to settling and the upward flux due to convection (Bartlett, 1969; Huppert and Sparks, 1980). This approach predicted that crystals stay in suspension if the settling velocity is smaller than the convective velocity.

However, this approach did not consider the fact that the convective velocities decrease near the bottom. The first systematic laboratory experiments on convective suspensions by Martin and Nokes (1988, 1989) showed that even when the settling velocity is much smaller than the convective velocity, the particles eventually settle down and they do so nearly as fast as in the absence of convection. They argued that because the convective velocity at the lower boundary vanishes, the particles cannot be reentrained and remain at the boundary. Yet, in one experiment, the amount of suspended particles remained constant after some partial initial sedimentation.

Tonks and Melosh (1990) addressed the problem of suspension in convective systems using the analogy with the entrainment of particles in shear flows. They argued that turbulence is the key factor for suspension for both shear flows and convection. According to their arguments, turbulence was not strong enough in Martin and Nokes' (1988, 1989) experiments. In particular, they argued that sedimentation in convective systems does not occur if the settling velocity is smaller than the effective friction velocity associated with turbulent fluctuations. However, Solomatov and Stevenson (1993a) pointed out that the Reynolds number in the only experiment where Martin and Nokes reported suspension was about 0.03, which is below the critical value for the transition to turbulence by two orders of magnitude. Thus, it remained unclear how particles are reentrained from the bottom.

Solomatov et al. (1993) investigated the mechanisms of entrainment experimentally using polystyrene spheres in an aqueous $CaCl_2$ solution. They showed that for either laminar or turbulent convection, the particles are moved at the bottom by the stresses generated by thermal plumes. These stresses are much larger than the stresses generated by turbulence. They also showed that the particles are not reentrained right away

but they are piled together in 'dunes.' If the flow is strong enough, it picks up the particles at the crests of the dunes and reentrains them back into the interior region of the convective layer. The critical condition for reentrainment is

$$\frac{\tau_T}{\Delta\rho g d} \sim 0.1 \quad [31]$$

where

$$\tau_T = \left(\frac{\eta_l \alpha g F}{c_p}\right)^{1/2} \quad [32]$$

is the convection stress scale and d is the diameter of the particles.

Application of these criteria to magma oceans is still not straightforward because the conditions in magma oceans are quite different from the laboratory experiments. Although the stress [32] dominates at low Rayleigh numbers, at high Rayleigh numbers, the situation can reverse. In fact, in the hard turbulence regime, the stresses associated with turbulence are larger than [32] (Shraiman and Siggia, 1990). The former are

$$\tau^* = \rho u^{*2} \quad [33]$$

where $u^* = u_0/x^*$ (eqn [28]). With typical values of $x^* \sim 60$ and $u_0 \sim 10\ \mathrm{m\,s^{-1}}$, one gets $\tau^* \sim 50\ \mathrm{Pa}$. Assuming that eqn [31] can be applied to the hard turbulence regime, that is,

$$\frac{\tau^*}{\Delta\rho g d} \sim 0.1 \quad [34]$$

one can get that crystals up to several centimeters in diameter can be suspended in magma oceans. The criterion [31] gives much smaller critical crystal size, by almost two orders of magnitude.

Perhaps a more important factor is that unlike non-interacting particles in the laboratory experiments, crystals can form solid bonds as observed in experiments on magmas (e.g., Lejeune and Richet, 1995). Formation and growth of solid bonds are a well-studied phenomenon in metallurgy, which plays a key role in liquid phase sintering (e.g., German, 1985). The bonds between crystals and the underlying bed can easily prevent reentrainment.

9.04.6.2 Energetics of Convective Suspensions

Even though reentrainment of crystals from the bottom of the convective layer is the key process that keeps particles suspended in laboratory experiments, this is not the most important factor that controls sedimentation in magma oceans. The crucial factor for magma oceans is that the presence of crystals can suppress convection as a result of viscous heating and density stratification associated with crystal settling (Solomatov and Stevenson, 1993a).

The total amount of energy released per unit time due to crystal settling is

$$\Phi = u_s g \phi \Delta\rho V \quad [35]$$

where u_s is the settling velocity and V is the volume of the magma ocean. Thus, convection has to do work to suspend crystals.

The total amount of energy per unit time, which is available for mechanical work, is

$$W \approx \frac{\alpha g L}{c_p} FA \approx FA \quad [36]$$

where A is the surface area of the Earth and $\alpha g L/c_p \sim 1$. Most of this work is spent to overcome viscous friction associated with convection, that is, just to keep convection going.

Experiments show that only a small fraction $\varepsilon \sim 0.1$–1% of W is available for reentrainment (Solomatov et al., 1993) If reentrainment is possible, the equilibrium volume fraction of crystals is determined by the condition that

$$\Phi = \varepsilon W \quad [37]$$

When $\Phi > \varepsilon W$, the sedimentation rate exceeds the reentrainment rate and vice versa.

When viscous dissipation due to sedimentation exceeds the available mechanical power, that is,

$$\Phi > W \quad [38]$$

convection is suppressed and suspension ceases to exist (Solomatov and Stevenson, 1993a). Although this theoretical prediction has not been tested experimentally, Solomatov and Stevenson (1993a) showed that this condition approximately coincides with the condition that the turbulence collapses due to density stratification caused by crystal settling (Solomatov and Stevenson, 1993a). The latter problem is analogous to the problem of turbulence in stratified fluids and has been well studied in laboratory experiments (Hopfinger, 1987).

When viscous dissipation exceeds the total heat loss rate from the magma ocean, that is,

$$\Phi > FA \quad [39]$$

then cooling can continue only after fractionation of some amount of crystals reduces Φ to the level where $\Phi < FA$.

9.04.6.3 Conditions for Fractional Crystallization

Fractional crystallization is expected to happen when the condition [39] is satisfied (cooling is impossible without sedimentation) and even when a less severe condition [38] is satisfied (convection is suppressed) although for deep magma oceans, these two conditions almost coincide. In either case, crystal–melt segregation occurs independently of whether or not there is any reentrainment of crystals at the bottom of the magma ocean.

To estimate the crystal size at which this happens, we need the equation for the settling velocity

$$u_s = f_\phi \frac{\Delta\rho g d^2}{18\eta_l} \quad [40]$$

where f_ϕ is a hindered settling function such that $f_\phi = 1$ at $\phi = 0$ (Davis and Acrivos, 1985). If the crystal fraction in the magma ocean varies from 0 to the maximum packing fraction $\phi_m = 0.6$, then the average crystal fraction is about $\phi \sim 30\%$ at which $f_\phi \sim 0.15$ (Davis and Acrivos, 1985).

The critical crystal size above which fractional crystallization occurs is then found from eqn [38]:

$$d_f = \left(\frac{18\alpha\eta_l FAL}{f_\phi g c_p \Delta\rho^2 \phi V}\right)^{1/2} \quad [41]$$

or

$$d_f \approx 10^{-3}\left(\frac{\eta_l}{0.1\,\text{Pa s}}\right)^{1/2}\left(\frac{F}{10^6\,\text{W m}^{-2}}\right)^{1/2}\text{m} \quad [42]$$

9.04.6.4 Conditions for Equilibrium Crystallization

A simple condition for the magma ocean to crystallize without any substantial chemical differentiation is one that requires that the sedimentation time is much smaller than the crystallization time (Solomatov, 2000).

The crystallization time is

$$t_c \approx \frac{(\Delta H\phi + c_p\Delta T)M}{FA} \approx 400\,\text{years} \quad [43]$$

where $\Delta T \sim 1000$ K is the average temperature drop upon crystallization of the magma ocean up to the crystal fraction $\phi \sim 60\%$ and M is the mass of the magma ocean.

The sedimentation time even in the presence of turbulent convection is (Martin and Nokes, 1988)

$$t_s \approx \frac{L}{u_s} \quad [44]$$

Crystallization is faster than sedimentation, that is, $t_c < t_s$, provided the crystal size is smaller than

$$d_e = \left(\frac{18L\eta_l}{f_\phi g\Delta\rho t_c}\right)^{1/2} \quad [45]$$

or

$$d_e \approx 10^{-3}\left(\frac{\eta_l}{0.1\,\text{Pa s}}\right)^{1/2}\left(\frac{F}{10^6\,\text{W m}^{-2}}\right)^{1/2}\text{m} \quad [46]$$

The difference between d_f and d_e is negligible compared to the uncertainties in the crystal size. Therefore, the critical crystal size that separates equilibrium and fractional crystallization is

$$d_{crit} \approx 10^{-3}\left(\frac{\eta_l}{0.1\,\text{Pa s}}\right)^{1/2}\left(\frac{F}{10^6\,\text{W m}^{-2}}\right)^{1/2}\text{m} \quad [47]$$

The fact that these two estimates are very close to each other is not surprising – their ratio scales approximately as

$$\frac{d_e}{d_f} \sim \left(\frac{\Delta\rho/\rho c_p}{\alpha T\Delta S}\right)^{1/2} \quad [48]$$

where we used $M \sim \rho AL$ and $\Delta H\phi + c_p\Delta T \sim \Delta H\phi \sim T\Delta S\phi$.

The entropy change upon melting ΔS is of the order of several k_B per atom (Stishov, 1988), while $c_p = 3k_B$ per atom. Therefore, $\Delta S \sim c_p$. Also, $\Delta\rho/\rho \sim \alpha T \sim 0.1$. Thus, $d_e/d_f \sim 1$.

9.04.7 Crystal Size in the Magma Ocean

The estimate of the critical crystal size $\sim$1 mm separating equilibrium and fractional crystallization is significantly smaller than the early estimates, which predicted that the critical crystal size is of the order of tens of meters (Miller et al., 1991a,b; Tonks and Melosh, 1990). If the critical crystal size were indeed this big, no physical mechanism could allow the crystals to reach this size during the lifetime of the magma ocean. The new estimate is in the range of typical crystal sizes observed in magmas (e.g., Cashman and Marsh, 1988), thus bringing up the question of whether the crystal size in the magma ocean was larger or smaller than $\sim$1 mm.

9.04.7.1 What Processes Control the Crystal Size in the Magma Ocean?

In thermodynamic equilibrium, the crystal diameter d is related to the crystal fraction ϕ and the number of crystals N per unit volume as (assuming spherical shape)

$$d = \left(\frac{6\phi}{\pi N}\right)^{1/3} \quad [49]$$

The number of crystals is controlled by nucleation in the descending flow. When crystal-free magma reaches the pressure where the temperature of the convecting magma ocean (adiabat) drops below the liquidus, crystals nucleate and grow until the equilibrium crystal fraction is reached.

If the number of crystals does not change with time, then the crystals grow simply because the equilibrium crystal fraction ϕ changes with depth along the adiabat – increases on the way down and decreases on the way up. However, dissolution of smaller crystals and growth of larger crystals decrease the number of crystals per unit volume and, thus, increase the average crystal size according to eqn [49]. This process is called Ostwald ripening.

9.04.7.2 Nucleation

When the temperature in the downwelling convective flow drops below the liquidus, crystals form via nucleation and growth mechanism (Figures 5 and 6). During nucleation, tiny crystals (tens of atoms in diameter, depending on the supercooling) precipitate from the supercooled melt. The

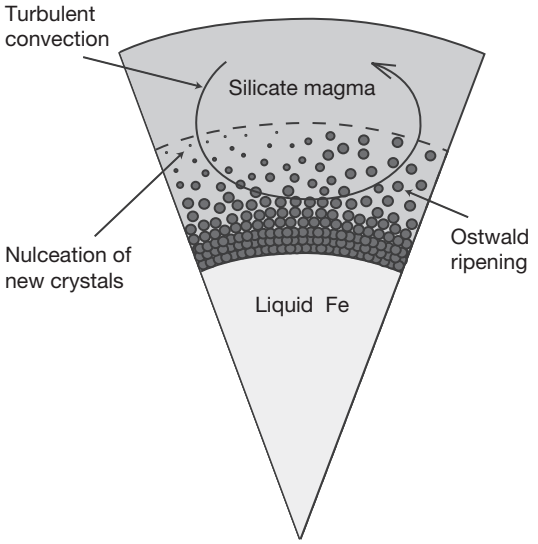


Figure 5 Magma ocean in the early stages of crystallization. Nucleation occurs in the downgoing convective flow when it enters the two-phase region. Crystals continue to grow because of Ostwald ripening – larger crystals grow at the expense of smaller ones. The crystals that have resided longer in the two-phase region are schematically shown bigger.

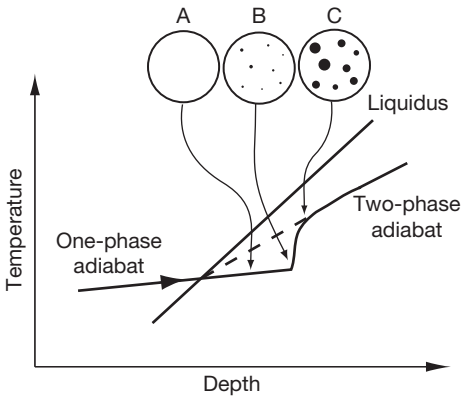


Figure 6 Evolution of a fluid parcel (large circle) in the downwelling flow. When the temperature of a fluid parcel drops below the liquidus, the temperature of the parcel initially follows a metastable one-phase adiabat (solid line) rather than the two-phase adiabat (dashed line), without any crystallization (stage A – pure liquid). When the supercooling reaches the nucleation barrier, an avalanche-like nucleation produces tiny crystals (stage B – liquid with nuclei). The subsequent crystal growth brings the temperature of the fluid up to the equilibrium two-phase adiabat (stage C – nearly equilibrium liquid–crystal mixture).

nucleation rate (the number of nuclei produced per unit volume per unit time) is an extremely sensitive function of the supercooling $\Delta T = T_{\mathrm{liq}} - T$ – the difference between the equilibrium value of the liquidus and the actual temperature of the fluid. It can approximately be described as

$$J(\Delta T) = a \exp\left(-\frac{B}{\Delta T^2}\right) \qquad [50]$$

$$B = \frac{16\pi\sigma_{\mathrm{app}}^3 T_0^2}{3k_{\mathrm{B}} T \rho^2 \Delta H^2} \qquad [51]$$

where a is a constant, T_0 is the melting temperature of the crystallizing phase, and σ_{app} is the apparent surface energy. Once a crystal has nucleated, it grows at a rate that can be controlled by either interface kinetics or diffusion. In the former case, the bottleneck in crystal growth is the attachment of atoms to the crystal (e.g., in some cases, the atom needs to find an imperfection on the crystal surface to be able to attach to it), while in the latter case, the attachment is relatively easy and the bottleneck is the long-range transport of atoms in the melt surrounding the growing crystal.

Although nucleation and crystal growth in the magma ocean might seem to be very difficult to constrain, fortunately, this process is mathematically very similar to bulk crystallization of continuously cooling liquids in laboratory conditions (Cashman, 1993; Flemings et al., 1976; Grove, 1990; Grove and Walker, 1977; Ichikawa et al., 1985; Lofgren et al., 1974; Smith et al., 1991; Walker et al., 1978). In both cases, the controlling parameter is the cooling rate $\dot{T}$. The only difference is that the effective cooling rate of an adiabatically descending fluid parcel is the rate of change of the difference between the liquidus temperature T_{liq} and the adiabatic temperature T_{ad}:

$$\dot{T} = u_0 \left(\frac{\mathrm{d}T_{\mathrm{liq}}}{\mathrm{d}z} - \frac{\mathrm{d}T_{\mathrm{ad}}}{\mathrm{d}z}\right) \qquad [52]$$

where $\mathrm{d}T_{\mathrm{liq}}/\mathrm{d}z$ is the liquidus gradient and $\mathrm{d}T_{\mathrm{ad}}/\mathrm{d}z$ is the adiabatic gradient.

Solomatov and Stevenson (1993c) solved the problem of nucleation in continuously cooling multicomponent liquids for interface kinetics controlled growth. Their solution is somewhat similar to the one obtained by Raizer (1960) (see also Zel'dovich and Raizer, 2002) for condensation in an adiabatically expanding cloud of vapor except that it uses a more general mathematical method developed by Buyevich and Mansurov (1991). Solomatov (1995b) solved this problem for the case when crystal growth is controlled by diffusion rather than interface kinetics and showed that this solution is more consistent with various experimental data on silicates and alloys.

According to Solomatov (1995b), nucleation in continuously cooling liquids operates during a very short period of time (an avalanche-like event). The abrupt start of nucleation is due to the fact that the liquid needs to reach a critical supersaturation (the nucleation barrier). The short duration and abrupt cessation of the nucleation are due to the fact that the precipitated crystals quickly change the composition of the fluid, which reduces the supersaturation below the nucleation barrier. Although the duration of nucleation is short, this is the key process that determines the number of crystals nucleated per unit volume, and thus, the size of these crystals after the system reaches an equilibrium.

The number of nuclei produced per unit volume during the nucleation period is calculated as follows:

$$N = 0.2 \frac{\zeta^{3/2} \dot{T}^{3/2} \theta^{3/2}}{B \left|\mathrm{d}T_{\mathrm{liq}}/\mathrm{d}\phi\right| D^{3/2}} \qquad [53]$$

where D is the diffusion coefficient and $\zeta = \Delta T/\Delta C$ determines the relationship between the supercooling ΔT and the supersaturation ΔC. It is of the order of the difference between liquidus and solidus temperatures, $\zeta \sim T_{\mathrm{liq}} - T_{\mathrm{sol}}$.

The parameter θ in eqn [53] is nearly constant ($\sim$30). It can be found from the transcendental equation

$$\theta = \ln\left[\frac{6.05\left|\mathrm{d}T_{\mathrm{liq}}/\mathrm{d}\phi\right| a B^{3/2} D^{3/2}}{\zeta^{3/2} \dot{T}^{5/2} \theta^3}\right] \qquad [54]$$

From eqns [49] and [53] with $\phi \sim 60\%$, we estimate

$$d_{\mathrm{nucl}} \approx 10^{-3} \left(\frac{\sigma_{\mathrm{app}}}{0.02\,\mathrm{J\,m^{-2}}}\right) \left(\frac{D}{10^{-9}\,\mathrm{m^2\,s^{-1}}}\right)^{1/2} \left(\frac{u_0}{10\,\mathrm{m\,s^{-1}}}\right)^{-1/2} \mathrm{m} \qquad [55]$$

This relationship can be interpreted as follows. The velocity controls the cooling rate of fluid parcels: the faster the cooling rate, the further the system is driven into the metastable state, the higher the nucleation rate and the more crystals are nucleated. This corresponds to the well-known fact that faster cooling produces smaller crystals. The diffusion coefficient controls the crystal size somewhat indirectly: during the nucleation and growth stage, the thermodynamic equilibrium is achieved via both the nucleation of new crystals and the growth of already nucleated crystals. The larger the diffusion coefficient, the more this balance is shifted toward growth and the fewer crystals are nucleated. The surface energy controls nucleation: the larger the surface energy, the more difficult it is to nucleate crystals and the fewer crystals are produced.

Note that the apparent cooling rates $\dot{T}$ in the magma ocean are not that much different from those in the laboratory

conditions (Figure 7). This means that there is no extrapolation in this parameter. The parameter that can significantly affect the estimate of the crystal size is the apparent surface energy, σ_{app}. It was argued that it is not the normal surface energy but rather an apparent one for nucleation (Dowty, 1980; Solomatov, 1995b). The value of 0.02 $J\,m^{-2}$ (Table 1) is based on the analogy with other cases that suggest that it is almost one order of magnitude smaller compared to the usual one. Higher values of σ_{app} in the magma ocean cannot be excluded. This means that it can increase the crystal size by as much as one order of magnitude.

It is also worth noting that because of a large value of the logarithm in eqn [54], the value of θ and thus, the crystal size, is insensitive to the uncertainties in the parameters under the logarithm. In particular, the most uncertain parameter is the prefactor a (the uncertainties can be as large as 10 orders of magnitude). However, this uncertainty effects the results only by a factor of 2 or so.

9.04.7.3 Ostwald Ripening

For diffusion-controlled Ostwald ripening, crystal size is calculated as follows (Lifshitz and Slyozov, 1961; Voorhees, 1992):

$$d_{ost}^3 - d_{nucl}^3 = \frac{32}{9} b_\phi \alpha_0 D t_{ost} \qquad [56]$$

where d_{nucl} is the initial crystal size after nucleation, $\alpha_0 = 2\sigma c_\infty v_m / RT$, σ is the surface energy, c_∞ is the equilibrium concentration of the crystallizing mineral in the melt, v_m is the molar volume of the crystallizing mineral, t_{ost} is the time available for Ostwald ripening, and b_ϕ is a function of ϕ such that $b_\phi = 1$ in the limit $\phi \ll 1$. The time available for Ostwald ripening is roughly the characteristic residence time in the two-phase region, that is, $t_{ost} \sim H/u_0$. Neglecting the initial crystal size d_{nucl}, we obtain that

$$d_{ost} \approx 10^{-3} \left(\frac{D}{10^{-9}\,m^2\,s^{-1}}\right)^{1/3} \left(\frac{u_0}{10\,m\,s^{-1}}\right)^{-1/3} m \qquad [57]$$

This is similar to estimate [55]. Thus, Ostwald ripening does not increase the crystal size substantially.

The previous estimates show that the crystal size during the early crystallization of the magma ocean is very close to the critical crystal size separating fractional and equilibrium crystallization of the magma ocean (Figure 8), which is about 1 mm. This means that both equilibrium and fractional crystallization (up to 60% crystal fraction) are equally acceptable within the uncertainties of the physical parameters.

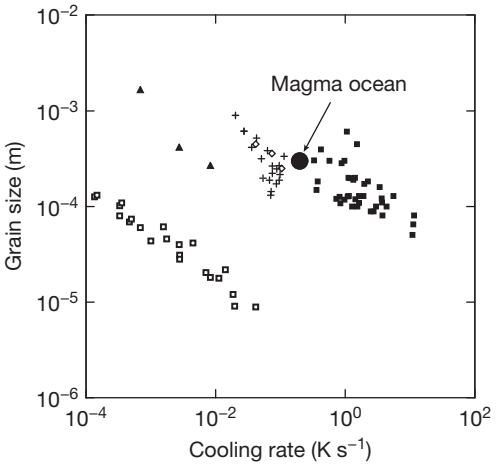


Figure 7 Crystal size versus cooling rate for five crystallizing phases (from Solomatov, 1995b): Sn (Sn–Pb system, solid boxes), Al (Al–Cu, pluses), Si (Al–Si, diamonds), diopside (diopside–plagioclase, solid triangles), and plagioclase (diopside–plagioclase, open boxes). The data are from Lofgren et al. (1974), Flemings et al. (1976), Grove and Walker (1977), Walker et al. (1978), Ichikawa et al. (1985), Grove (1990), Smith et al. (1991), and Cashman (1993). The data approximately follow the theoretical 1/2 slope. The location of the magma ocean is shown.

9.04.8 Crystallization Beyond the Rheological Transition

As was mentioned earlier, at the crystal fraction around $\phi_m \sim 60\%$, crystal–melt mixture undergoes a rheological transition to a solid-like behavior, and the deformation is controlled by the solid-state creep of crystals. Solomatov and Stevenson (1993b) and Solomatov (2000) showed that rapid cooling and crystallization may continue even at $\phi > \phi_m$.

This can be explained as follows. When the viscosity of magma abruptly increases from liquid-like viscosity to the solid-like viscosity around $\phi \sim \phi_m$, convection stops at the bottom of the magma ocean or somewhere near it – the hot core may keep the base of the mantle molten even after crystallization of the magma ocean. A nonconvecting layer with the crystal fraction $\phi \sim \phi_m$ starts growing from the bottom of the magma ocean (Figure 9). If solid-state convection does not start, then in about 400 years, eqn [43], the temperature in the entire magma ocean would approximately follow the curve $\phi \sim \phi_m = \text{const}$. However, the temperature gradient of the magma ocean, which is partially crystallized to $\phi \sim \phi_m$, is

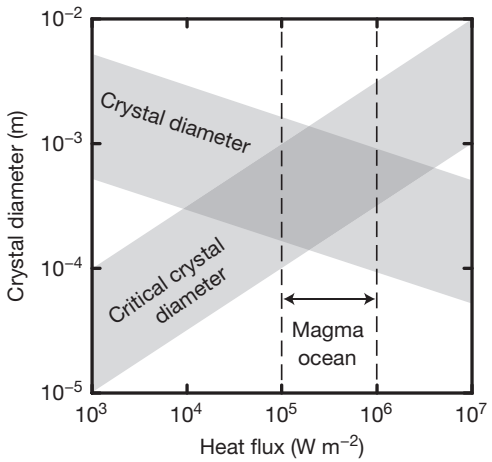


Figure 8 The crystal size during the early stages of crystallization of the magma ocean (assuming that it is controlled by nucleation in the downwelling flow) and the critical crystal size for suspension are shown as functions of the heat flux. The width of the curves (gray bands) represents the uncertainty range of one order of magnitude.

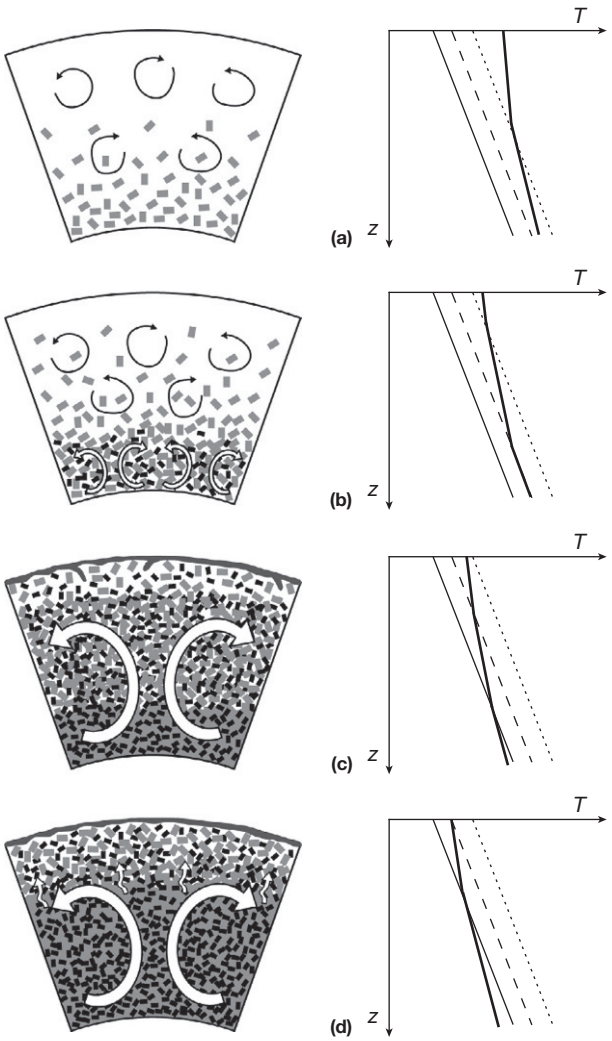


Figure 9 Crystallization of the magma ocean: (a) The lower part of the magma ocean is below liquidus (dotted line), convection is controlled by melt viscosity, and the temperature (heavy solid line) is adiabatic; (b) high-viscosity, gravitationally unstable region with the maximum packing crystal fraction forms near the bottom of the magma ocean ($\phi = \phi_m$ curve is shown with a dashed line); (c) cooling beyond $\phi = \phi_m$ proceeds via solid-state convection, which is still fast at this stage. The temperature in the high-viscosity region below solidus (solid line) can be superadiabatic; (d) $\phi > \phi_m$ everywhere and the rate of cooling and crystallization of the remaining melt are controlled by solid-state convection and melt percolation. The gray and black crystals schematically illustrate different mineral phases.

roughly parallel to solidus and thus is steeper than the adiabat. Since the adiabat is the temperature profile of a neutrally stable fluid, a superadiabatic temperature profile would make the mantle gravitationally unstable.

The instability is so fast that one can ignore thermal diffusion and treat it like a Rayleigh–Taylor instability. The driving effective density contrast is of the order of

$$\Delta\rho_{RT} \approx \frac{1}{2}\alpha\rho L\left(\frac{dT_{sol}}{dz} - \frac{dT_{ad}}{dz}\right) \approx 100\,\mathrm{kg\,m^{-3}} \qquad [58]$$

for the whole mantle ($L \approx 3 \times 10^6$ m).

The timescale for the overturn can then be estimated as (Turcotte and Schubert, 2002)

$$t_{RT} \approx 26\frac{\eta_s}{\Delta\rho_{RT} g L} \qquad [59]$$

where η_s is the viscosity of the solid mantle and g is the gravity.

The viscosity η_s of the lower mantle is poorly constrained. The present-day lower mantle viscosity is around 10^{22} Pa s (e.g., King, 1995). The temperature of the mantle is much below solidus, by ~1000–1500 K at the base of the mantle (Boehler, 1996, 2000; Zerr et al., 1998). Thus, the viscosity near the solidus must be substantially lower than 10^{22} Pa s. Based on experimental and theoretical constraints on the lower mantle viscosity (Béjina et al., 2003; Ita and Cohen, 1998; Karato and Li, 1992; Li et al., 1996; Wright and Price, 1993; Yamazaki and Karato, 2001; Yamazaki et al., 2000), we can assume that the viscosity near the solidus is about 10^{18} Pa s for the grain size of 10^{-3} m.

The presence of melt reduces the viscosity further. Equation [18] suggests that the viscosity is ~10^4–10^5 times smaller at $\phi \sim \phi_m$ and one order of magnitude smaller when only 10% melt remains in the mantle. These estimates are very approximate because eqn [18] is a phenomenological equation whose underlying physical mechanisms are poorly understood and thus the errors due to extrapolation to the magma ocean conditions are unknown. Nevertheless, eqns [18] and [59] predict that for a near-solidus mantle, the overturn is of the order of 100 years, and for $\phi \sim \phi_m$, it can be as short as a few days.

Thus, the instability starts developing soon after the maximum packing fraction is reached at the bottom of the magma ocean and takes the form of solid-state convection, which is the primary mechanism of cooling and crystallization in the range $\phi_m < \phi < 1$. A complete solidification front would follow the rheological front toward the surface (Figure 9). When the adiabatic temperature in the solidified regions becomes so low that the viscosity exceeds about 10^{19} Pa s, then Rayleigh–Taylor will be unable to catch up with the solidification front. It is likely that the temperatures in the high-viscosity regions will be superadiabatic. In this case, a substantial superadiabatic temperature gradient can be created upon crystallization of the magma ocean. The instability would continue after crystallization of the magma ocean. It may take a form of overturn as suggested for Mars by Elkins-Tanton et al. (2005), which would create a subadiabatic and gravitationally stable mantle. It should be emphasized that this remains one of the most difficult and uncertain aspects of magma ocean crystallization.

9.04.9 The Last Stages of Crystallization

9.04.9.1 Cessation of Suspension

The first important transition occurs when the molten layer disappears, that is, the potential temperature drops below liquidus everywhere. The crystal size increases because the sequence 'nucleation–growth–dissolution' changes to just 'growth.' During the nucleation–growth–dissolution cycle, the time available for Ostwald ripening is controlled by the residence time of crystals in the partially molten region, which is of the order of $t_{ost} \sim 3H/u_0 \sim 10^6$ s or roughly 1 week. When the completely molten layer disappears, crystals never exit the

two-phase region and the characteristic time for crystal growth is much larger. It can approximately be estimated as the time it takes for the potential temperature to drop from the liquidus temperature to the critical temperature for the rheological transition (i.e., while liquid-state convection still occurs). Since the difference between liquidus T_{liq} and solidus T_{sol} is about 600 K near the surface (Figure 1) and $\phi \approx 60\,\%$ is reached approximately in the middle between liquidus and solidus (Figure 2), then $t'_{ost} \sim 3 \times 10^9$ s (100 years) according to eqn [43]. The crystal size increases by $(t'_{ost}/t_{ost})^{1/3} \sim 14$ times, that is, $d'_{ost} \sim 10^{-2}$ m. If crystal–melt segregation did not start earlier, at this stage, it is impossible to avoid (Figure 8).

In addition, the surface temperature would eventually drop below 1500 K (Figure 4). If substantial amounts of water were present at that time, the atmosphere would change from silicate to steam one. The blanketing effect of the steam atmosphere can reduce the heat flux to 10^2–10^3 W m^{-2} (Abe and Matsui, 1986; Kasting, 1988; Zahnle et al., 1988). This would increase the crystal size even further and at the same time substantially reduce the ability of convection to suspend crystals (Figure 8).

The pressure at the bottom of the partially molten layer (the complete crystallization front) at the time when the completely molten layer disappears can be estimated from an adiabat, which starts at the liquidus temperature at the surface. The bottom of the partially molten layer is the pressure where this adiabat intersects with the solidus. It depends strongly on the details of the thermodynamics of the partially molten layer. Calculations by Miller et al. (1991a,b), Solomatov and Stevenson (1993b), Abe (1997), and Solomatov (2000) (Figures 1 and 3) suggest that the bottom of the partially molten layer is around 30–40 GPa.

When crystal–melt segregation begins, convection can become quite complex. This regime is poorly understood. It may develop layers and instabilities as experiments on convective suspension show (e.g., Koyaguchi et al., 1993; Sparks et al., 1993). Although convection is suppressed by sedimentation, it may not necessarily cease. Crystal fraction near the surface is below 60%, and thus, the viscosity in a thin region below the surface is still small. This might be sufficient to maintain convection at least in this layer. Note that convection is driven by the instabilities of the 1 cm thick surface thermal boundary layer and the heat flux does not depend much on the processes at large depth, such as crystal–melt segregation. This means that as long as a low-viscosity layer exists below the surface, one might expect that this layer would continue to convect vigorously generating the high heat flux predicted by eqns [19] and [21].

9.04.9.2 Formation of Thin Crust Within the Thermal Boundary Layer

Even though a low-viscosity region may exist below the surface, a stable crust can form within the thin thermal boundary layer, which would reduce the heat flux. Indeed, when the potential temperature drops below the liquidus, the surface temperature can reach the critical one where the crystal fraction is about 60% and can even reach the solidus (Figures 1, 2, and 4). First, it seems that it is unlikely that a solid or partially solid crust can form on the timescale of $\delta^2/4\kappa \sim 30$ s – this is too short for crystals to nucleate or grow. Second, if a rigid crust does form on top of the thermal boundary layer, its effect on the heat flux is not as big as one might expect. This can be estimated as follows. The viscosity decreases within the thermal boundary layer due to both temperature and crystal fraction. As a result, while the whole boundary layer cannot participate in convection because of the rigid crust, the bottom part which has the lowest viscosity can. This convection regime is called stagnant lid convection.

The heat flux in the stagnant lid convection regime can approximately be described by eqns [19] and [21], provided the temperature difference $T_m - T_s$ is replaced by the rheological temperature scale (Davaille and Jaupart, 1993a,b; Fowler, 1985; Morris and Canright, 1984; Solomatov, 1995a; Solomatov and Moresi, 2000). For magma, it is calculated as follows (Davaille and Jaupart, 1993b; Reese and Solomatov, 2006):

$$\Delta T_{rh} \approx 2\left|\frac{\mathrm{d}\ln\eta}{\mathrm{d}T}\right|^{-1}_{T=T_m} = 2\left|\frac{\partial\ln\eta}{\partial T} + \frac{\partial\ln\eta}{\partial\phi}\frac{\mathrm{d}\phi}{\mathrm{d}T}\right|^{-1}_{T=T_m} \quad [60]$$

Assuming an Arrhenius-dependence of melt viscosity on temperature, $\eta_l \sim \exp(E_l/RT)$, where the activation energy $E_l = 190$ kJ mol^{-1} (Liebske et al., 2005b) and using eqn [17] and Figure 2, we obtain that near liquidus $\Delta T_{rh} \approx 100$ K. This scale is mostly determined by the variation in the crystal fraction within the boundary layer.

Thus, the driving temperature difference for convection can be reduced by a factor of 2–6 compared to the temperature difference without solid crust (Figure 4) and the heat flux can drop by one order of magnitude (eqns [19] and [21]).

9.04.9.3 Cessation of Liquid-State Convection

When the crystal fraction increases to 60% all the way to the surface, the effective viscosity of magma is controlled by solid-state deformation of crystals. Convection becomes much slower, the heat flux drops, and a thick crust forms at the surface. The location of the bottom of the remaining partially molten layer can be obtained from an adiabat starting at the critical temperature for the rheological transition. Figure 1 suggests that such an adiabat intersects the solidus around 10 GPa or 300 km. This estimate is rather uncertain since even a 100 K error in the estimate of the rheological transition increases this pressure by a factor of 2. A similar factor can be caused by variations in the solidus and in the thermodynamics of partial melt. Also, this gives only a lower limit because the temperatures can be somewhat superadiabatic due to the strong variation of the viscosity with depth and because of the compositional stratification at shallow depth. After solid-state convection replaces liquid-state convection, percolation of melt through the solid matrix becomes an important process.

9.04.9.4 Percolation

Melt extraction from the remaining partially molten layer can be controlled by either the percolation of melt through the solid matrix or the compaction of the solid matrix (Bercovici et al., 2001; McKenzie, 1984). The compaction length is of the order of meters, which is much smaller than the depth L' of the remaining partially molten layer. This implies that compaction does not effect melt extraction and the rate-limiting process is

melt percolation. Although the melt fraction and the viscosities of solid and liquid vary in the partially molten region, an order-of-magnitude melt percolation time is

$$t_{\text{diff}} \sim \frac{L'}{u_{\text{perc}}} \quad [61]$$

where

$$u_{\text{perc}} = \frac{g\Delta\rho d^2 \phi_{\text{l}}^2}{150\eta_{\text{l}}(1-\phi_{\text{l}})} \quad [62]$$

is the percolation velocity (Dullien, 1979; Soo, 1967), which is related to Darcy velocity, $u_{\text{D}} \approx \phi_{\text{l}} u_{\text{perc}}$, and $\phi_{\text{l}} = 1 - \phi$ is the melt fraction. Melt is likely to migrate very quickly at the initial stages of percolation (near the rheological transition), when the melt fraction is large and the melt viscosity is low, and significantly slow down near the solidus when only a few percent melt is left. For $\phi_{\text{l}} \sim 2\%$, $d \sim 1\,\text{mm}$ and $\eta_{\text{l}} \sim 100\,\text{Pa s}$ (the value of the viscosity of a low-pressure, high silica, polymerized magma just above the solidus, Kushiro, 1980, 1986),

$$t_{\text{diff}} \sim 10^8 \left(\frac{\phi_{\text{l}}}{0.02}^2\right) \left(\frac{d}{10^{-3}\,\text{m}}\right)^2 \left(\frac{\eta_{\text{l}}}{100\,\text{Pa s}}\right) \text{years} \quad [63]$$

9.04.9.5 Crystal–Melt Density Inversions

Crystal–melt density inversions in the magma ocean (Agee, 1998) might result in the formation of more than one molten layer. This is very important because it determines the direction of melt migration and chemical differentiation in the magma ocean. It is now realized that stable molten layers could have been formed not only near the surface but also in the deep mantle, in the upper mantle, and at the core–mantle boundary (Agee, 2008; Akins et al., 2004; de Koker et al., 2013; Matsukage et al., 2005; Murakami and Bass, 2011; Nomura et al., 2011; Ohtani, 2009; Ohtani and Maeda, 2001; Stixrude and Karki, 2005). Constraining these density inversions is quite challenging and the results from different groups sometimes differ (e.g., Andrault et al., 2012). Besides the experimental difficulties, there are intrinsic difficulties in determining density inversions because of their high sensitivity to composition and, in particular, to poorly constrained iron and water contents in the melt.

Formation and evolution of a dense magmatic layer at the base of the mantle early in the Earth's history would have important consequences for the Earth's evolution (Labrosse et al., 2007). This layer can produce a hidden geochemical reservoir, which can be a host to various incompatible elements. For example, it can explain the apparent deficit of ^{40}K and the low degassing rates of ^{40}Ar and the difference in ^{142}Nd/^{144}Nd ratios between chondrites and the Earth's mantle. It can also explain He and Ne signatures of ocean island basalts (Coltice et al., 2011).

9.04.9.6 Remelting due to Melt Extraction

The extraction of melt is accompanied by gravitational energy release, which is converted to heat via viscous dissipation associated with melt migration. Assuming that most of the energy goes to melting, Solomatov and Stevenson (1993b) showed that the additional amount of melt generated is proportional to

$$R_{\text{m}} = \frac{\Delta\rho g L'}{\rho \Delta H} \quad [64]$$

The value of R_{m} is only about 0.1 for 50 km deep magma chambers but close to unity for a layer of $L' \sim 500\,\text{km}$ depth. Thus, the total degree of melting can roughly be twice as much as the initial degree of melting before melt extraction.

9.04.9.7 Solid-State Convection

The timescale for complete crystallization of the mantle depends on the style of convection in the remaining partially molten layers. When the deformation is controlled by solid-state creep, the heat flux is determined by the equations, which are similar to those describing plate tectonics. In its simplest form, it can be written as (Davies, 1990; Turcotte and Schubert, 2002)

$$F = 0.3k(T_{\text{m}} - T_{\text{s}})^{4/3} \left(\frac{\alpha\rho g}{\kappa\eta_{\text{s}}}\right)^{1/3} \quad [65]$$

For the viscosity $\eta_{\text{s}} \sim 10^{18}\,\text{Pa s}$, and the temperature difference across the surface thermal boundary layer $T_{\text{m}} - T_{\text{s}} \sim 1500\,\text{K}$ (i.e., the surface temperature is around 300 K), the earlier equation gives $F \sim 1\,\text{W m}^{-2}$. The thickness of the thermal boundary layer is about 1 km. The crystallization of a partially molten upper layer of thickness $L' \sim 500\,\text{km}$ takes

$$t_{\text{conv}} \sim \frac{\Delta H(1-\phi_{\text{m}})\rho L'}{F} \sim 10\,\text{Myr} \quad [66]$$

If one assumes that the surface layer is rigid and convection occurs in the stagnant lid regime, then the driving temperature $T_{\text{m}} - T_{\text{s}} \approx 1500\,\text{K}$ in eqn [65] needs to be replaced by the rheological temperature scale eqn [60]. Equations [60] and [18] give $\Delta T_{\text{rh}} \approx 100\,\text{K}$, where we assumed $\text{d}\phi/\text{d}T \approx 5 \times 10^{-4}\,\text{K}^{-1}$ near the rheological transition (the melt fraction is about 40%; **Figure 2**) and $E \approx 240\,\text{kJ mol}^{-1}$, which is the activation energy for diffusion creep in water-saturated olivine (Karato and Wu, 1993). The heat flux can drop by a factor of ~ 40, thus increasing the crystallization time to $\sim 400\,\text{Myr}$. If the viscosity η_{s} and/or $\text{d}\phi/\text{d}T$ is larger (the latter depends on the details of the phase diagram), the crystallization time can increase further, perhaps, to 2 billion years or so.

It is also possible that density stratification due to chemical differentiation completely suppresses convection. In the stagnant lid convection regime, where the driving temperature differences are small, this does not take much (Zaranek and Parmentier, 2004). If this happens, cooling and crystallization would be controlled by thermal diffusion whose timescale $\sim L'^2/4\kappa$ is about 2 billion years. These order-of-magnitude estimates show the crystallization time can be comparable with the age of the Earth. This means that it is possible that the Earth never had a chance to crystallize completely.

The early studies inevitably considered a simple, one-dimensional model of magma ocean crystallization in which any chemical heterogeneities would form laterally homogeneous layers. However, the mantle may acquire lateral heterogeneities as well, for example, because of an asymmetrical pattern

of impact-induced melting and incomplete mixing. Smaller, yet still sufficiently large impacts capable of melting a part of the mantle may produce localized magma oceans, which form lateral heterogeneities (Reese et al., 2007b). These heterogeneities are likely to be not only chemical but also rheological and would be difficult to mix by mantle convection later in the Earth's history (Solomatov and Reese, 2008). A growing number of isotopic evidence do indeed suggest that early differentiation of the Earth was heterogeneous and these heterogeneities may have never fully mixed by mantle convection (Bennett et al., 2008; Caro, 2011; Touboul et al., 2012).

9.04.10 Magma Oceans on Other Planets

The magma ocean hypothesis is becoming an important building block in the narrative of the formation and evolution of not only the Earth but also other planets (Albarède and Blichert-Toft, 2007; Brown and Elkins-Tanton, 2009; Elkins-Tanton, 2012; Elkins-Tanton et al., 2005; Reese and Solomatov, 2006; Reese et al., 2007a; Sleep, 2000; Stevenson, 2008; Zahnle et al., 2007).

Venus is similar in size to Earth and is likely to have undergone some form of widespread melting. The extent of melting and the evolution of the Venusian magmasphere can be very different from the Earth because of the stochastic nature of the formation process. The Venusian magmasphere may have lasted longer and could have even expanded during the Venusian history due to radioactive heating and inefficient cooling in the absence of plate tectonics (Reese et al., 2007a). This may explain the extensive volcanism in the past billion years or so and the existence of recent (within a few million years) volcanism (Smrekar et al., 2010).

Mars is roughly half the size of the Earth and seems to have differentiated fast. The estimates range from around a few tens of millions of years (Foley et al., 2005; Kleine et al., 2002) to a few millions of years (Dauphas and Pourmand, 2011). The small size of Mars and the isotopic data suggest more localized magma oceans leading to a more heterogeneous planet (Reese and Solomatov, 2006; Reese et al., 2007b). This type of magma oceans is in contrast to a global, more or less uniform magma ocean usually assumed for the Earth.

Mercury has a peculiar structure characterized by an extremely thin mantle. A giant impact has been proposed as one of the possible causes (Benz et al., 1988). More recently, the magma ocean hypothesis has been invoked to explain the observed low-iron content of the crust by a postcrystallization mantle overturn (Brown and Elkins-Tanton, 2009).

Although Io, Jupiter's satellite, has been known for its extraordinary level of volcanism caused by tidal heating, the possibility that Io may have a global magma ocean has been suggested only recently (Khurana et al., 2011). A partially molten layer with thickness of about 50 km and 20% melt seems to explain best the magnetic data returned by the Galileo mission.

Isotopic data suggest that even small bodies underwent melting and differentiation early during planetary formation. High-precision Hf–W data for iron meteorites suggest that this happened in about 1–1.5 million years (Kruijer et al., 2012; Qin et al., 2008; Scherstén et al., 2006). It is important to note that the estimates of differentiation times assume equilibrium melting. Impact-induced metal/silicate segregation may have occurred under nonequilibrium conditions (Tomkins et al., 2013). If so, this implies that the time might be longer than the equilibrium model suggest and that melting may have been caused by impacts rather than by ^{26}Al.

The growing list of planets outside of the solar system now includes planets that are currently likely to be at least partially molten. For example, CoRoT-7b, about 500 light years away from Earth and about 1.7 times the size of Earth, is estimated to be melted due to tidal heating and its proximity to the star, which keeps its surface temperature around 1800–2600 K (Léger et al., 2009). A more recently discovered Kepler-10b, which is 560 light years away from Earth and is closer in size to Earth (about 1.4 times larger), also has a very hot surface, around 1800 K.

9.04.11 Summary

Crystallization of the molten Earth was undoubtedly very complex. A variety of physical and chemical processes determined the final product of crystallization during the first 10^8 years, before the long-term solid-state convection took over. This includes convection in liquid, solid, and partially molten states, crystal settling during the early stages of crystallization, and extraction of residual melt from the partially molten mantle during the later stages.

A scenario that seems to agree with the geochemical constraints and is physically feasible is as follows. The latest and the largest impact melts a significant part of the Earth. Gravitational instabilities quickly redistribute the material so that denser materials (due to temperature and composition) accumulate at the bottom and lighter ones accumulate at the top. A small fraction of the mantle that survived the impact and remained solid sink to the bottom of the magma ocean.

In the beginning, the viscosity of the magma ocean is very low ($\sim$0.1 Pa s), the magma ocean is vigorously convecting, and the heat flux $\sim 10^6$ W m^{-2}. The temperature profile is approximately adiabatic. Since the melting curve is likely to be steeper than the adiabat, crystallization proceeds from the bottom up. It takes about $\sim 10^3$ years to crystallize the lower part of the mantle. Highly turbulent convection helps to cool and crystallize the magma ocean to about 60% crystal fraction (the rheological transition from a low-viscosity suspension to a high-viscosity partially molten solid).

Whether or not fractional crystallization is prevented during the early stages of crystallization depends on the crystal size. It is controlled by nucleation in the downgoing convective flow and by Ostwald ripening. The crystal diameter established by these processes can be sufficiently small ($\sim$1 mm) to allow equilibrium crystallization. The lower part of the mantle underwent only a small degree of crystal–melt segregation, which can be reconciled with the geochemical constrained on the fractionation of minor and trace elements.

Crystallization beyond the crystal fraction of 60% is accomplished by solid-state creep convection. Although the viscosity of solids is very high ($\sim 10^{18}$ Pa s at the solidus and $\sim 10^{14}$ Pa s near 60% crystal fraction), convection remains sufficiently fast and is able to help crystallize the lower mantle completely.

Several important changes happen during the last stages of crystallization of the magma ocean. The purely molten region disappears, and the crystals undergo continuous growth rather

than cycle between nucleation, growth, and dissolution. The early silicate atmosphere changes to a steam atmosphere whose blanketing effect can reduce the heat flux by several orders of magnitude. Convection becomes weaker, while the crystals grow larger and cannot be suspended by convection. Crystal settling or flotation (depending on pressure) creates a compositional and density stratification, which tends to suppress convection although vigorous convection and cooling are likely to continue near the surface. When the crystal fraction reaches the maximum packing fraction of about $\sim$60% all the way to the surface, the heat flux drops to $\sim 1\ \mathrm{W\,m^{-2}}$ because of the viscosity jump from liquid-like to solid-like viscosity (the rheological transition). Formation of the solid crust can reduce the heat flux further by one order of magnitude. All these changes contribute to the onset of crystal–melt segregation in the magma ocean. The depth of the differentiated region is from a few hundred kilometers to 1000 km, with the melt fraction varying from zero at the bottom to $\sim$60% at the top.

Crystallization of this remaining partially molten layer takes 10^7–10^9 years, depending on the regime of mantle convection (surface recycling or stagnant lid convection). This is comparable with the time it takes the melt to escape via percolation. Radiogenic heating and subsolidus convection become the main factors determining the subsequent evolution of the planet.

The two very different timescales, 10^3 years for crystallization of the lower mantle and 10^7–10^9 years for crystallization of the upper mantle, suggest that whenever the Earth was melted by a giant impact, the lower mantle healed very quickly, while the upper mantle never had enough time to crystallize completely. If so, then iron delivered by impacts accumulated at the base of the upper mantle before it sank into the Earth's core. Thus, chemical equilibrium between iron and silicates was established at high pressures near the bottom of the partially molten upper mantle in agreement with models of abundances of siderophile elements.

Although the model of equilibrium crystallization provides a simple conceptual framework that can help to reconcile various geochemical constraints, some form of fractional crystallization cannot be excluded within the uncertainties of physical parameters. Also, fractional crystallization is inevitable at some stage of crystallization even within this model. The dynamics of fractional crystallization is undoubtedly very complex and involves a variety of dynamic processes including crystal settling/flotation, convective mixing in the melt and in the partially molten solid, melt percolation, gravitational instabilities due to unstable density stratification, and compositional and convective layering. This is an important problem that needs to be addressed in the future. Finally, future studies will need to consider primordial differentiation of the Earth's mantle in combination with planetary accretion, core formation, and atmospheric evolution.

References

Abe Y (1993) Physical state of the very early Earth. *Lithos* 30: 223–235.

Abe Y (1995) Early evolution of the terrestrial planets. *Journal of Physics of the Earth* 43: 515–532.

Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.

Abe Y (2011) Protoatmospheres and surface environment of protoplanets. *Earth, Moon, and Planets* 108: 9–14.

Abe Y and Matsui T (1986) Early evolution of the Earth: Accretion, atmosphere formation, and thermal history. *Journal of Geophysical Research* 91: E291–E302.

Abe Y, Ohtani E, Okuchi T, Righter K, and Drake M (2000) Water in the early Earth. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 413–433. Tucson, AZ: University of Arizona Press.

Adjaoud O, Steinle-Neumann G, and Jahn S (2012) Transport properties of Mg_2SiO_4 liquid at high pressure: Physical state of a magma ocean. *Earth and Planetary Science Letters* 312: 463–470.

Agee CB (1990) A new look at differentiation of the Earth from melting experiments on the Allende meteorite. *Nature* 346: 834–837.

Agee CB (1998) Crystal-liquid density inversions in terrestrial and lunar magmas. *Physics of the Earth and Planetary Interiors* 107: 63–74.

Agee CB (2008) Static compression of hydrous silicate melt and the effect of water on planetary differentiation. *Earth and Planetary Science Letters* 265: 641–654.

Agee CB and Walker D (1988) Mass balance and phase density constraints on early differentiation of chondritic mantle. *Earth and Planetary Science Letters* 90: 144–156.

Ahrens TJ (1992) A magma ocean and the Earth's internal water budget. In: Agee CB and Longhi J (eds.) *Workshop on the Physics and Chemistry of Magma Oceans from 1 bar to 4 Mbar, LPI Tech. Rep. 92-03*, pp. 13–14. Houston: Lunar and Planetary Institute.

Akins JA, Luo SN, Asimow PD, and Ahrens TJ (2004) Shock-induced melting of $MgSiO_3$ perovskite and implications for melts in Earth's lowermost mantle. *Geophysical Research Letters* 31. http://dx.doi.org/10.1029/2004GL020237.

Albarède F and Blichert-Toft J (2007) The split fate of the early Earth, Mars, Venus, and Moon. *Comptes Rendus Geoscience* 339: 917–927.

Albarède F, Blichert-Toft J, Vervoort JD, Gleason JD, and Rosing M (2000) Hf–Nd isotope evidence for a transient dynamic regime in the early terrestrial mantle. *Nature* 404: 488–490.

Amelin Y, Lee DC, Halliday AN, and Pidgeon RT (1999) Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons. *Nature* 399: 252–255.

Andrade ENC (1952) Viscosity of liquids. *Proceedings of the Royal Society of London, Series A* 215: 36–43.

Andrault D, Petitgirard S, Lo Nigro G, et al. (2012) Solid-liquid iron partitioning in Earth's deep mantle. *Nature* 487: 354–359.

Arzi AA (1978) Critical phenomena in the rheology of partially melted rocks. *Tectonophysics* 44: 173–184.

Asimow PD, Hirschmann MM, and Stolper EM (1997) An analysis of variations in isentropic melt productivity. *Philosophical Transactions of the Royal Society of London, Series A* 355: 255–281.

Bartlett RB (1969) Magma convection, temperature distribution, and differentiation. *American Journal of Science* 267: 1067–1082.

Béjina F, Jaoul O, and Liebermann RC (2003) Diffusion in minerals at high pressure: A review. *Physics of the Earth and Planetary Interiors* 139: 3–20.

Bennett VC, Brandon AD, and Nutman AP (2008) Coupled ^{142}Nd–^{143}Nd isotopic evidence for Hadean mantle dynamics. *Science* 318: 1907–1910.

Bennett VC, Nutman AP, and McCulloch MT (1993) Nd isotopic evidence for transient highly-depleted mantle reservoirs in the early history of the Earth. *Earth and Planetary Science Letters* 119: 299–317.

Benz W and Cameron AGW (1990) Terrestrial effects of the giant impact. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 61–67. New York: Oxford University Press.

Benz W, Cameron AGW, and Melosh HJ (1989) The origin of the Moon and the single impact hypothesis, III. *Icarus* 81: 113–131.

Benz W, Slattery WL, and Cameron AGW (1986) The origin of the Moon and the single impact hypothesis, I. *Icarus* 66: 515–535.

Benz W, Slattery WL, and Cameron AGW (1987) The origin of the Moon and the single impact hypothesis, II. *Icarus* 71: 30–45.

Benz W, Slattery WL, and Cameron AGW (1988) Collisional stripping of Mercury's mantle. *Icarus* 74: 516–528.

Bercovici D, Ricard Y, and Schubert G (2001) A two-phase model for compaction and damage 1. General theory. *Journal of Geophysical Research* 106: 8887–8906.

Bizzarro M, Baker JA, Haack H, Ulfbeck D, and Rosing MG (2003) Early history of Earth's crust–mantle system inferred from hafnium isotopes in chondrites. *Nature* 421: 931–933.

Boehler R (1992) Melting of the Fe–FeO and the Fe–FeS systems at high pressure: Constraints on core temperatures. *Earth and Planetary Science Letters* 111: 217–227.

Boehler R (1996) Melting temperature of the Earth's mantle and core: Earth's thermal structure. *Annual Review of Earth and Planetary Sciences* 24: 15–40.

Boehler R (2000) High-pressure experiments and the phase diagram of lower mantle and core materials. *Reviews of Geophysics* 38: 221–245.

Bottinga Y, Richet P, and Sipp A (1995) Viscosity regimes of homogeneous silicate melts. *American Mineralogist* 80: 305–318.

Bottinga Y and Weill DF (1972) The viscosity of magmatic silicate liquids: A model for calculation. *American Journal of Science* 272: 438–475.

Boubnov BM and Golitsyn GS (1986) Experimental study of convective structures in rotating fluids. *Journal of Fluid Mechanics* 167: 503–531.

Boubnov BM and Golitsyn GS (1990) Temperature and velocity field regimes of convective motions in a rotating plane fluid layer. *Journal of Fluid Mechanics* 219: 215–239.

Bowring SA and Housh T (1995) The Earth's early evolution. *Science* 269: 1535–1540.

Boyet M and Carlson RW (2005) ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.

Brinkman HC (1952) The viscosity of concentrated suspensions and solutions. *Journal of Chemical Physics* 20: 571–581.

Brown S and Elkins-Tanton LT (2009) Compositions of Mercury's earliest crust from magma ocean models. *Earth and Planetary Science Letters* 286: 446–455.

Buyevich YA and Mansurov VV (1991) Kinetics of the intermediate stage of phase transition in batch crystallization. *Journal of Crystal Growth* 104: 861–867.

Cameron AGW (1997) The origin of the Moon and the single impact hypothesis V. *Icarus* 126: 126–137.

Campbell GA and Forgacs G (1990) Viscosity of concentrated suspensions: An approach based on percolation theory. *Physical Review A* 41: 4570–4573.

Canup RM (2004) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.

Canup RM (2008) Accretion of the Earth. *Philosophical Transactions of the Royal Society A* 366: 4061–4075.

Canup RM and Agnor CB (2000) Accretion of the terrestrial planets and the Earth–Moon system. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 113–129. Tucson, AZ: University of Arizona Press.

Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.

Canup RM and Esposito LW (1996) Accretion of the Moon from an impact-generated disk. *Icarus* 119: 427–446.

Canuto VM and Dubovikov MS (1998) Two scaling regimes for rotating Rayleigh–Benard convection. *Physical Review Letters* 80: 281–284.

Caro G (2011) Early silicate Earth differentiation. *Annual Review of Earth and Planetary Sciences* 39: 31–58.

Caro G, Bourdon B, Birck JL, and Moorbath S (2003) ^{146}Sm–^{142}Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle. *Nature* 423: 428–432.

Caro G, Bourdon B, Wood BJ, and Corgne A (2005) Trace-element fractionation in Hadean mantle generated by melt segregation from a magma ocean. *Nature* 436: 246–249.

Cashman KV (1993) Relationship between plagioclase crystallization and cooling rate in basaltic melts. *Contributions to Mineralogy and Petrology* 113: 126–142.

Cashman KV and Marsh BD (1988) Crystal size distribution (CSD) in rocks and the kinetics and dynamics of crystallization II. Makaopuhi lava lake. *Contributions to Mineralogy and Petrology* 99: 292–305.

Castaing B, Gunaratne G, Heslot F, et al. (1989) Scaling of hard thermal turbulence in Rayleigh–Benard convection. *Journal of Fluid Mechanics* 204: 1–30.

Caughey SJ and Palmer SG (1979) Some aspects of turbulence structure through the depth of the convective boundary layer. *Quarterly Journal of the Royal Meteorological Society* 105: 811–827.

Chambers JE and Wetherill GW (1998) Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus* 136: 304–327.

Chen R, Fernando HJS, and Boyer DL (1989) Formation of isolated vortices in a rotating convecting fluid. *Journal of Geophysical Research* 94: 18445–18453.

Clark SP (1957) Radiative transfer in the Earth's mantle. *Transactions of the American Geophysical Union* 38: 931–938.

Collerson KD, Campbell LM, Weaver BL, and Palacz ZA (1991) Evidence for extreme mantle fractionation in early Archaean ultramafic rocks from northern Labrador. *Nature* 349: 209–214.

Coltice N, Moreira M, Hernlund J, and Labrosse S (2011) Crystallization of a basal magma ocean recorded by helium and neon. *Earth and Planetary Science Letters* 308: 193–199.

Cooper RF and Kohlstedt DL (1986) Rheology and structure of olivine-basalt partial melts. *Journal of Geophysical Research* 91: 9315–9323.

Corgne A, Liebske C, Wood BJ, Rubie DC, and Frost DJ (2005) Silicate perovskite-melt partitioning of trace elements and geochemical signature of a deep perovskitic reservoir. *Geochimica et Cosmochimica Acta* 69: 485–496.

Corgne A and Wood BJ (2002) $CaSiO_3$ and $CaTiO_3$ perovskite-melt partitioning of trace elements: Implications for gross mantle differentiation. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2001GL014398.

Costa A (2005) Viscosity of high crystal content melts: Dependence on solid fraction. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2005GL024303.

Dauphas N and Pourmand A (2011) Hf-W-Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* 473: 489–493.

Davaille A and Jaupart C (1993a) Transient high Rayleigh number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.

Davaille A and Jaupart C (1993b) Thermal convection in lava lakes. *Geophysical Research Letters* 20: 1827–1830.

Davies GF (1990) Heat and mass transport in the early Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 175–194. New York: Oxford University Press.

Davis RH and Acrivos A (1985) Sedimentation of noncolloidal particles at low Reynolds numbers. *Annual Review of Fluid Mechanics* 17: 91–118.

de Koker N, Karki BB, and Stixrude L (2013) Thermodynamics of the MgO–SiO_2 liquid system in Earth's lowermost mantle from first principles. *Earth and Planetary Science Letters* 361: 58–63.

de Koker N and Stixrude L (2009) Self-consistent thermodynamic description of silicate liquids, with application to shock melting of MgO periclase and $MgSiO_3$ perovskite. *Geophysical Journal International* 178: 162–179.

Deardorff JW (1970) Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection. *Journal of Atmospheric Science* 27: 1211–1213.

Dingwell DB, Courtial P, Giordano D, and Nichols ARL (2004) Viscosity of peridotite liquid. *Earth and Planetary Science Letters* 226: 127–138.

Dowty E (1980) Crystal growth and nucleation theory and the numerical simulation of igneous crystallization. In: Hargraves RV (ed.) *Physics of Magmatic Processes*, pp. 419–485. Princeton: Princeton University Press.

Drake MJ (2000) Accretion and primary differentiation of the Earth: A personal journey. *Geochimica et Cosmochimica Acta* 64: 2363–2370.

Dullien FAL (1979) *Porous Media: Fluid Transport and Pore Structure*. San Diego, CA: Academic Press.

Einstein A (1906) Eine neue Bestimmung der Moleküldimensionen. *Annalen der Physik* 19: 289–306.

Elkins-Tanton LT (2011) Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth and Planetary Science Letters* 271: 181–191.

Elkins-Tanton LT (2012) Magma oceans in the inner Solar System. *Annual Review of Earth and Planetary Sciences* 40: 113–139.

Elkins-Tanton LT, Zaranek SE, Parmentier EM, and Hess PC (2005) Early magnetic field and crust on Mars from magma ocean cumulate overturn. *Earth and Planetary Science Letters* 236: 1–12.

Elsasser WM (1963) Early history of the Earth; dedicated to F. G. Houtermans on his sixtieth birthday. In: Geiss J and Goldberg ED (eds.) *Earth Science and Meteoritics*, pp. 1–30. Amsterdam: North-Holland.

Fernando HJS, Chen RR, and Boyer DL (1991) Effects of rotation on convective turbulence. *Journal of Fluid Mechanics* 228: 513–547.

Flasar FM and Birch F (1973) Energetics of core formation: A correction. *Journal of Geophysical Research* 78: 6101–6103.

Flemings MC, Riek RG, and Young KP (1976) Rheocasting. *Materials Science and Engineering* 25: 103–117.

Foley CN, Wadhwa M, Borg LE, Janney PE, Hines R, and Grove TL (2005) The early differentiation history of Mars from W-182–Nd-142 isotope systematics in the SNC meteorites. *Geochimica et Cosmochimica Acta* 69: 4557–4571.

Fowler AC (1985) Fast thermoviscous convection. *Studies in Applied Mathematics* 72: 189–219.

Frankel NA and Acrivos A (1967) On the viscosity of a concentrated suspension of solid spheres. *Chemical Engineering Science* 22: 847–853.

Gans RF (1972) Viscosity of the Earth's core. *Journal of Geophysical Research* 77: 360–366.

Gasparik T and Drake MJ (1995) Partitioning of elements among two silicate perovskites, superphase B, and volatile-bearing melt at 23 GPa and 1500–1600 °C. *Earth and Planetary Science Letters* 134: 307–318.

German RM (1985) *Liquid Phase Sintering*. New York: Plenum Press.

Ghiorso MS (1997) Thermodynamic models of igneous processes. *Annual Review of Earth and Planetary Sciences* 25: 221–241.

Glazier JA (1999) Evidence against 'ultrahard' thermal turbulence at very high Rayleigh numbers. *Nature* 398: 307–310.

Golitsyn GS (1980) Geostrophic convection. *Doklady Akademii Nauk SSSR* 251: 1356–1360.

Golitsyn GS (1981) Structure of convection in rapid rotation. *Doklady Akademii Nauk SSSR* 261: 317–320.

Goody RM (1995) *Principles of Atmospheric Physics and Chemistry*. New York: Oxford University Press.

Grossmann S and Lohse D (1992) Scaling in hard turbulent Rayleigh–Bènard flow. *Physical Review A* 46: 903–917.
Grove TL (1990) Cooling histories of lavas from Serocki volcano. In: *Proceedings of the Ocean Drilling Program, 106/109,* pp. 3–8.
Grove TL and Walker D (1977) Cooling histories of Apollo 15 quartz-normative basalts. In: *Proceedings of the 8th Lunar and Planetary Science Conference*, pp. 1501–1520.
Halliday AN (2004) Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* 427: 505–509.
Halliday A, Rehkämper M, Lee D-C, and Yi W (1996) Early evolution of the Earth and Moon: New constraints from Hf–W isotope geochemistry. *Earth and Planetary Science Letters* 142: 75–89.
Harrison TM, Blichert-Toft J, Muller W, Albarède F, Holden P, and Mojzsis SJ (2005) Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* 310: 1947–1950.
Hayashi C, Nakazawa K, and Mizuno H (1979) Earth's melting due to the blanketing effect of the primordial dense atmosphere. *Earth and Planetary Science Letters* 43: 22–28.
Herzberg C and Gasparik T (1991) Garnet and pyroxenes in the mantle: A test of the majorite fractionation hypothesis. *Journal of Geophysical Research* 96: 16263–16274.
Hirose K, Shimizu N, van Westrenen W, and Fei Y (2004) Trace element partitioning in Earth's lower mantle and implications for geochemical consequences of partial melting at the core–mantle boundary. *Physics of the Earth and Planetary Interiors* 146: 249–260.
Hirth G and Kohlstedt DL (1995a) Experimental constraints on the dynamics of the partially molten upper mantle: Deformation in the diffusion creep regime. *Journal of Geophysical Research* 100: 1981–2001.
Hirth G and Kohlstedt DL (1995b) Experimental constraints on the dynamics of the partially molten upper mantle 2. Deformation in the dislocation creep regime. *Journal of Geophysical Research* 100: 15441–15449.
Holland KG and Ahrens TJ (1997) Melting of $(Mg,Fe)_2SiO_4$ at the core–mantle boundary of the Earth. *Science* 275: 1623–1625.
Hopfinger EJ (1987) Turbulence in stratified fluids: A review. *Journal of Geophysical Research* 92: 5287–5303.
Hopfinger EJ (1989) Turbulence and vortices in rotating fluids. In: Germain P, Piau M, and Caillerie D (eds.) *Theoretical and Applied Mechanics*, pp. 117–138. New York: Elsevier.
Hopfinger EJ, Browand FK, and Gagne Y (1982) Turbulence and waves in a rotating tank. *Journal of Fluid Mechanics* 125: 505–534.
Hostetler CJ and Drake MJ (1980) On the early global melting of the terrestrial planets. In: *Proceedings of the 11th Lunar and Planetary Science Conference,* pp. 1915–1929.
Huppert HE and Sparks RSJ (1980) The fluid dynamics of a basaltic magma chamber replenished by influx of hot, dense ultrabasic magma. *Contributions to Mineralogy and Petrology* 75: 279–289.
Ichikawa K, Kinoshita Y, and Shimamura S (1985) Grain refinement in Al–Cu binary alloys by rheocasting. *Transactions of the Japan Institute of Metals* 26: 513–522.
Ida S, Canup RM, and Stewart GR (1997) Lunar accretion from an impact-generated disk. *Nature* 389: 353–357.
Ita J and Cohen RE (1998) Diffusion in MgO at high pressure: Implications for lower mantle rheology. *Geophysical Research Letters* 25: 1095–1098.
Ito E, Kubo A, Katsura T, and Walter MJ (2004) Melting experiments of mantle materials under lower mantle conditions with implications for magma ocean differentiation. *Physics of the Earth and Planetary Interiors* 143: 397–406.
Ito E and Takahashi E (1987) Melting of peridotite at uppermost lower-mantle conditions. *Nature* 328: 514–517.
Jacobsen SB (2005) The Hf–W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences* 33: 531–570.
Jin Z-M, Green HW, and Zhou Y (1994) Melt topology in partially molten mantle peridotite during ductile deformation. *Nature* 372: 164–167.
Kadanoff LP (2001) Turbulent heat flow: Structures and scalings. *Physics Today* 54: 34–39.
Karato S-I and Li P (1992) Diffusion creep in perovskite: Implications for the rheology of the lower mantle. *Science* 255: 1238–1240.
Karato S-I and Murthy VR (1997a) Core formation and chemical equilibrium in the Earth – I. Physical considerations. *Physics of the Earth and Planetary Interiors* 100: 61–79.
Karato S-I and Murthy VR (1997b) Core formation and chemical equilibrium in the Earth – II. Chemical consequences for the mantle and core. *Physics of the Earth and Planetary Interiors* 100: 81–95.
Karato S-I and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.
Karki BB and Stixrude LP (2010) Viscosity of $MgSiO_3$ liquid at Earth's mantle conditions: Implications for an early magma ocean. *Science* 328: 740–742.
Kasting JF (1988) Runaway and moist greenhouse atmosphere and the evolution of Earth and Venus. *Icarus* 74: 472–494.
Kato T, Ringwood AE, and Irifune T (1988a) Experimental determination of element partitioning between silicate perovskites. garnets and liquids: Constraints on early differentiation of the mantle. *Earth and Planetary Science Letters* 89: 123–145.
Kato T, Ringwood AE, and Irifune T (1988b) Constraints on element partition coefficients between $MgSiO_3$ perovskite and liquid determined by direct measurements. *Earth and Planetary Science Letters* 90: 65–68.
Kaula WM (1979) Thermal evolution of earth and moon growing by planetesimals impacts. *Journal of Geophysical Research* 84: 999–1008.
Ke Y and Solomatov VS (2006) Early transient superplumes and the origin of the Martian crustal dichotomy. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JE002631.
Ke Y and Solomatov VS (2009) Coupled core–mantle thermal evolution of early Mars. *Journal of Geophysical Research* 114. http://dx.doi.org/10.1029/2008JE003291.
Khurana KK, Jia XZ, Kivelson MG, et al. (2011) Evidence of a global magma ocean in Io's interior. *Science* 332: 1186–1189.
King SD (1995) Models of mantle viscosity. In: Ahrens TJ (ed.) *Mineral Physics and Crystallography: A Handbook of Physical Constants*, pp. 227–236. Washington, DC: American Geophysical Union.
Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* 418: 952–955.
Knittle E and Jeanloz R (1989) Melting curve of (Mg, Fe)SiO_3 perovskite to 96 GPa: Evidence for a structural transition in lower mantle melts. *Geophysical Research Letters* 16: 421–424.
Kohlstedt DL and Zimmerman ME (1996) Rheology of partially molten mantle rocks. *Annual Review of Earth and Planetary Sciences* 24: 41–62.
Koyaguchi T, Hallworth MA, and Huppert HE (1993) An experimental study of the effects of phenocrysts on convection in magmas. *Journal of Volcanology and Geothermal Research* 55: 15–32.
Kraichnan RH (1962) Turbulent thermal convection at arbitrary Prandtl number. *Physics of Fluids* 5: 1374–1389.
Krieger IM and Dougherty TJ (1959) A mechanism for non-Newtonian flow in suspensions of rigid spheres. *Transactions of the Society of Rheology* 3: 137–152.
Kruijer TS, Sprung P, Kleine T, et al. (2012) Hf-W chronometry of core formation in planetesimals inferred from weakly irradiated iron meteorites. *Geochimica et Cosmochimica Acta* 99: 287–304.
Kushiro I (1980) Viscosity, density, and structure of silicate melts at high pressures, and their petrological applications. In: Hargraves RB (ed.) *Physics of Magmatic Processes*, pp. 93–120. Princeton: Princeton University Press.
Kushiro I (1986) Viscosity of partial melts in the upper mantle. *Journal of Geophysical Research* 91: 9343–9350.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450: 866–869.
Lee D-C and Halliday AN (1995) Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* 378: 771–774.
Léger A, Rouan D, Schneider J, et al. (2009) Transiting exoplanets from the CoRoT space mission. *Astronomy & Astrophysics* 506: 287–302.
Lejeune A-M and Richet P (1995) Rheology of crystal-bearing silicate melts: An experimental study at high viscosities. *Journal of Geophysical Research* 100: 4215–4229.
Li J and Agee CB (1996) Geochemistry of mantle–core differentiation at high pressure. *Nature* 381: 686–689.
Li P, Karato S, and Wang ZC (1996) High-temperature creep in fine-grained polycrystalline $CaTiO_3$, an analogue material of $(Mg,Fe)SiO_3$. *Physics of the Earth and Planetary Interiors* 95: 19–36.
Liebske C, Corgne A, Frost DJ, Rubie DC, and Wood BJ (2005a) Compositional effects on element partitioning between Mg-silicate perovskite and silicate melts. *Contributions to Mineralogy and Petrology* 149: 113–128.
Liebske C, Schmickler B, Terasaki H, et al. (2005b) Viscosity of peridotite liquid up to 13 GPa: Implications for magma ocean viscosities. *Earth and Planetary Science Letters* 240: 589–604.
Lifshitz IM and Slyozov VV (1961) The kinetics of precipitation from supersaturated solid solution. *Journal of Physics and Chemistry of Solids* 19: 35–50.
Lofgren G, Donaldson CH, Williams RJ, Mullins O, and Usselman TM (1974) Experimentally reproduced textures and mineral chemistry of Apollo 15 quartz-normative basalts. In: *Proceedings of the 5th Lunar Science Conference,* pp. 549–567.
Longhi J (1980) A model of early lunar differentiation. In: *Proceedings of the 11th Lunar Science Conference,* pp. 289–315.

Marcq E (2012) A simple 1-D radiative-convective atmospheric model designed for integration into coupled models of magma ocean planets. *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2011JE003912.

Marsh BD and Maxey MR (1985) On the distribution and separation of crystals in convecting magma. *Journal of Volcanology and Geothermal Research* 24: 95–150.

Martin D and Nokes R (1988) Crystal settling in a vigorously convecting magma chamber. *Nature* 332: 534–536.

Martin D and Nokes R (1989) A fluid dynamical study of crystal settling in convecting magmas. *Journal of Petrology* 30: 1471–1500.

Matsui T and Abe Y (1986) Formation of a "magma ocean" on the terrestrial planets due to the blanketing effect of an impact-induced atmosphere. *Earth, Moon, and Planets* 34: 223–230.

Matsukage KN, Jing Z, and Karato S-I (2005) Density of hydrous silicate melt at the conditions of Earth's deep upper mantle. *Nature* 438: 488–491.

McBirney AR and Murase T (1984) Rheological properties of magmas. *Annual Review of Earth and Planetary Sciences* 12: 337–357.

McFarlane EA and Drake MJ (1990) Element partitioning and the early thermal history of the Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 135–150. New York: Oxford University Press.

McFarlane EA, Drake MJ, and Rubie DC (1994) Element partitioning between Mg-perovskite, magnesiowustite, and silicate melt at conditions of the Earth's mantle. *Geochimica et Coscochimica Acta* 58: 5161–5172.

McGeary RK (1961) Mechanical packing of spherical particles. *Journal of the American Ceramic Society* 44: 513–522.

McKenzie D (1984) The generation and compaction of partially molten rock. *Journal of Petrology* 25: 713–765.

McKenzie D and Bickle MJ (1988) The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology* 29: 625–679.

Mei S, Bai W, Hiraga T, and Kohlstedt DL (2002) Influence of melt on the creep behavior of olivine-basalt aggregates under hydrous conditions. *Earth and Planetary Science Letters* 201: 491–507.

Mei S and Kohlstedt DL (2000a) Influence of water on plastic deformation of olivine aggregates 1. Diffusion creep regime. *Journal Geophysical Research* 105: 21457–21469.

Mei S and Kohlstedt DL (2000b) Influence of water on plastic deformation of olivine aggregates 2. Dislocation creep regime. *Journal of Geophysical Research* 105: 21471–21481.

Melosh HJ (1990) Giant impacts and the thermal state of the early Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 69–83. New York: Oxford University Press.

Miller GH, Stolper EM, and Ahrens TJ (1991a) The equation of state of a molten komatiite, 1. Shock wave compression to 36 GPa. *Journal of Geophysical Research* 96: 11831–11848.

Miller GH, Stolper EM, and Ahrens TJ (1991b) The equation of state of a molten komatiite, 2, Application to komatiite petrogenesis and the Hadean mantle. *Journal of Geophysical Research* 96: 11849–11864.

Mooney M (1951) The viscosity of a concentrated suspension of spherical particles. *Journal of Colloid Science* 6: 162–170.

Morris S and Canright D (1984) A boundary layer analysis of Bénard convection in a fluid of strongly temperature dependent viscosity. *Physics of the Earth and Planetary Interiors* 36: 355–373.

Mosenfelder JL, Asimow PD, Frost DJ, Rubie DC, and Ahrens TJ (2009) The $MgSiO_3$ system at high pressure: Thermodynamic properties of perovskite, postperovskite, and melt from global inversion of shock and static compression data. *Journal of Geophysical Research* 114. http://dx.doi.org/10.1029/2008JB005900.

Mostefaoui S, Lugmair GW, and Hoppe P (2005) Fe-60: A heat source for planetary differentiation from a nearby supernova explosion. *Astrophysical Journal* 625: 271–277.

Murakami M and Bass JD (2011) Evidence of denser $MgSiO_3$ glass above 133 gigapascal (GPa) and implications for remnants of ultradense silicate melt from a deep magma ocean. *Proceedings of the National Academy of Sciences* 108: 17286–17289.

Murase T and McBirney AR (1973) Properties of some common igneous rocks and their melts at high temperatures. *Geological Society of America Bulletin* 84: 3563–3593.

Murray JD (1965) On the mathematics of fluidization, Part I. Fundamental equations and wave propagation. *Journal of Fluid Mechanics* 21: 465–493.

Nakazawa K, Mizuno H, Sekiya M, and Hayashi C (1985) Structure of the primordial atmosphere surrounding the early Earth. *Journal of Geomagnetism and Geoelectricity* 37: 781–799.

Newsom HE and Taylor SR (1989) Geochemical implications of the formation of the Moon by a single giant impact. *Nature* 338: 29–34.

Nicolas A and Ildefonse B (1996) Flow mechanism and viscosity in basaltic magma chambers. *Geophysical Research Letters* 23: 2013–2016.

Nicolis G and Prigogine I (1977) *Self-Organization in Non-Equilibrium Systems*. New York: Wiley.

Niemela JJ, Skrbek L, Sreenivasan KR, and Donnelly RJ (2000) Turbulent convection at very high Rayleigh numbers. *Nature* 404: 837–840.

Nomura R, Ozawa H, Tateno S, et al. (2011) Spin crossover and iron-rich silicate melt in the Earth's deep mantle. *Nature* 473: 199–203.

Ohtani E (1985) The primordial terrestrial magma ocean and its implication for stratification of the mantle. *Physics of the Earth and Planetary Interiors* 38: 70–80.

Ohtani E (2009) Melting relations and the equation of state of magmas at high pressure: Application to geodynamics. *Chemical Geology* 265: 279–288.

Ohtani E and Maeda M (2001) Density of basaltic melt at high pressure and stability of the melt at the base of the lower mantle. *Physics of the Earth and Planetary Interiors* 193: 69–75.

Ohtani E and Sawamoto H (1987) Melting experiment on a model chondritic mantle composition at 25 GPa. *Geophysical Research Letters* 14: 733–736.

Persikov ES, Zharikov VA, Bukhtiyarov PG, and Polskoy SF (1997) The effect of volatiles on the properties of magmatic melts. *European Journal of Mineralogy* 2: 621–642.

Pierazzo E, Vickery AM, and Melosh HJ (1997) A reevaluation of impact melt production. *Icarus* 127: 408–423.

Pinkerton H and Sparks RSJ (1978) Field measurements of the rheology of lava. *Nature* 276: 383–385.

Poe BT, McMillan PF, Rubie DC, Chakraborty S, Yarger J, and Diefenbacher J (1997) Silicon and oxygen self-diffusivities in silicate liquids measured to 15 gigapascals and 2800 kelvin. *Science* 276: 1245–1248.

Presnall DC, Weng Y-H, Milholland CS, and Walter MJ (1998) Liquidus phase relations in the system $MgO–MgSiO_3$ at pressures up to 25 GPa – Constraints on crystallization of a molten Hadean mantle. *Physics of the Earth and Planetary Interiors* 107: 83–95.

Priestly CHB (1959) *Turbulent Transfer in the Lower Atmosphere*. Chicago, IL: University of Chicago Press.

Qin LP, Dauphas N, Wadhwa M, et al. (2008) Rapid accretion and differentiation of iron meteorite parent bodies inferred from 182Hf-182W chronometry and thermal modeling. *Earth and Planetary Science Letters* 273: 94–104.

Raizer Yu P (1960) Condensation of a cloud of vaporized matter expanding in vacuum. *Soviet Physics JETP (English Translation)* 10: 1229–1235.

Reese CC and Solomatov VS (2006) Fluid dynamics of local martian magma oceans. *Icarus* 184: 102–120.

Reese CC, Solomatov VS, and Orth CP (2007a) Mechanisms for cessation of magmatic resurfacing on Venus. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JE002782.

Reese CC, Solomatov VS, and Orth CP (2007b) Interaction between local magma ocean evolution and mantle dynamics on Mars. In: Foulger GR and Jurdy DM (eds.) *Plates, Plumes, and Planetary Processes*, pp. 913–932. Geological Society of America.

Righter K (2003) Metal-silicate partitioning of siderophile elements and core formation in the early Earth. *Annual Review of Earth and Planetary Sciences* 31: 135–174.

Righter K (2011) Prediction of metal–silicate partition coefficients for siderophile elements: An update and assessment of PT conditions for metal–silicate equilibrium during accretion of the Earth. *Earth and Planetary Science Letters* 304: 158–167.

Righter K and Drake MJ (1997) Metal-silicate equilibrium in a homogeneously accreting Earth: New results for Re. *Earth and Planetary Science Letters* 146: 541–553.

Righter K, Drake MJ, and Yaxley G (1997) Prediction of siderophile element metal-silicate partition coefficients to 20 GPa and 2800 °C: The effects of pressure, temperature, oxygen fugacity, and silicate and metallic melt compositions. *Physics of the Earth and Planetary Interiors* 100: 115–134.

Ringwood AE (1990) Earliest history of the Earth–Moon system. In: Newsom NE and Jones JH (eds.) *Origin of the Earth*, pp. 101–134. New York: Oxford University Press.

Ringwood AE and Kesson SE (1976) A dynamic model for mare basalt petrogenesis. In: *Proceedings of the 7th Lunar Planetary Science Conference*, pp. 1697–1722.

Robson GR (1967) Thickness of Etnean lavas. *Nature* 216: 251–252.

Roscoe R (1952) The viscosity of suspensions of rigid spheres. *British Journal of Applied Physics* 3: 267–269.

Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

Rudman M (1992) Two-phase natural convection: Implications for crystal settling in magma chambers. *Physics of the Earth and Planetary Interiors* 72: 153–172.

Rutter EH and Neumann DHK (1995) Experimental deformation of partially molten Westerly granite under fluid-absent conditions, with implications for the extraction of granitic magmas. *Journal of Geophysical Research* 100: 15697–15715.

Ryerson FJ, Weed HC, and Piwinskii AJ (1988) Rheology of subliquidus magmas: I. Picritic compositions. *Journal of Geophysical Research* 93: 3421–3436.
Saar MO, Manga M, Cashman KV, and Fremouw S (2001) Numerical models of the onset of yield strength in crystal-melt suspensions. *Earth Planet Science Letters* 187: 367–379.
Safronov VS (1964) The primary inhomogeneities of the Earth's mantle. *Tectonophysics* 1: 217–221.
Safronov VS (1978) The heating of the Earth during its formation. *Icarus* 33: 3–12.
Scarfe CM and Takahashi E (1986) Melting of garnet peridotite to 13 GPa and the early history of the upper mantle. *Nature* 322: 354–356.
Scherstén A, Elliott T, Hawkesworth C, et al. (2006) Hf-W evidence for rapid differentiation of iron meteorite parent bodies. *Earth and Planetary Science Letters* 241: 530–542.
Sears WD (1993) Tidal dissipation and the giant impact origin for the Moon. *Lunar and Planetary Science Conference* 23: 1255–1256.
Shankland TJ, Nitsan U, and Duba AG (1979) Optical absorption and radiative heat transport in olivine at high temperatures. *Journal of Geophysical Research* 84: 1603–1610.
Shaw HR (1969) Rheology of basalt in the melting range. *Journal of Petrology* 10: 510–535.
Shaw HR (1972) Viscosities of magmatic silicate liquids: An empirical method of prediction. *American Journal of Science* 272: 870–893.
Shaw HR, Wright TL, Peck DL, and Okamura R (1968) The viscosity of basaltic magma: An analysis of field measurements in Makaopuhi lava lake, Hawaii. *American Journal of Science* 266: 225–264.
Shraiman BI and Siggia ED (1990) Heat transport in high-Rayleigh-number convection. *Physical Review A* 42: 3650–3653.
Shukolyukov A and Lugmair GW (1993) Live iron-60 in the early solar system. *Science* 259: 1138–1142.
Siebert J, Badro J, Antonangeli D, and Ryerson FJ (2012) Metal silicate partitioning of Ni and Co in a deep magma ocean. *Earth and Planetary Science Letters* 321–322: 189–197.
Siggia ED (1994) High Rayleigh number convection. *Annual Review of Fluid Mechanics* 26: 137–168.
Sleep NH (2000) Evolution of the mode of convection within terrestrial planets. *Journal of Geophysical Research* 105. http://dx.doi.org/10.1029/2000JE001240.
Smith DM, Eady JA, Hogan LM, and Irwin DW (1991) Crystallization of a faceted primary phase in a stirred slurry. *Metallurgical Transactions A* 22: 575–584.
Smrekar SE, Stofan ER, Mueller N, et al. (2010) Recent hotspot volcanism on Venus from VIRTIS emissivity data. *Science* 328: 605–608.
Solomatov VS (1995a) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Solomatov VS (1995b) Batch crystallization under continuous cooling: Analytical solution for diffusion limited crystal growth. *Journal of Crystal Growth* 148: 421–431.
Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tucson, AZ: University of Arizona Press.
Solomatov VS and Moresi L-N (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on the Earth and other terrestrial planets. *Journal of Geophysical Research* 105: 21795–21818.
Solomatov VS, Olson P, and Stevenson DJ (1993) Entrainment from a bed of particles by thermal convection. *Earth and Planetary Science Letters* 120: 387–393.
Solomatov VS and Reese CC (2008) Grain size variations in the Earth's mantle and the evolution of primordial chemical heterogeneities. *Journal of Geophysical Research* 113. http://dx.doi.org/10.1029/2007JB005319.
Solomatov VS and Stevenson DJ (1993a) Suspension in convective layers and style of differentiation of a terrestrial magma ocean. *Journal of Geophysical Research* 98: 5375–5390.
Solomatov VS and Stevenson DJ (1993b) Nonfractional crystallization of a terrestrial magma ocean. *Journal of Geophysical Research* 98: 5391–5406.
Solomatov VS and Stevenson DJ (1993c) Kinetics of crystal growth in a terrestrial magma ocean. *Journal of Geophysical Research* 98: 5407–5418.
Sonett CP, Colburn DS, and Schwartz K (1968) Electrical heating of meteorite parent bodies and planets by dynamo induction from a pre-main sequence T Tauri solar wind. *Nature* 219: 924–926.
Soo SL (1967) *Fluid Dynamics of Multiphase Systems*. Waltham: Blaisdell.
Sparks RS, Huppert HE, Koyaguchi T, and Hallworth MA (1993) Origin of modal and rhythmic igneous layering by sedimentation in a convecting magma chamber. *Nature* 361: 246–249.
Srinivasan G, Goswami JN, and Bhandari N (1999) ^{26}Al in eucrite Piplia Kalan: Plausible heat source and formation chronology. *Science* 284: 1348–1350.
Stevenson DJ (1981) Models of the Earth's core. *Science* 214: 611–619.
Stevenson DJ (1987) Origin of the Moon – The collision hypothesis. *Annual Review of Earth and Planetary Sciences* 15: 271–315.
Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 231–249. New York: Oxford University Press.
Stevenson DJ (2008) A planetary perspective on the deep Earth. *Nature* 45: 261–265.
Stishov SM (1988) Entropy, disorder, melting. *Soviet Physics Uspekhi* 31: 52–67.
Stixrude L and Karki B (2005) Structure and freezing of $MgSiO_3$ liquid in Earth's lower mantle. *Science* 14: 297–299.
Thomas CW, Liu Q, Agee CB, Asimow PD, and Lange RA (2012) Multi-technique equation of state for Fe_2SiO_4 melt and the density of Fe-bearing silicate melts from 0 to 161 GPa. *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2012JB009403.
Thompson C and Stevenson DJ (1988) Gravitational instabilities in 2-phase disks and the origin of the Moon. *Astrophysical Journal* 333: 452–481.
Thomson W (1864) On the secular cooling of the Earth. *Transactions of Royal Society of Edinburgh* 23: 157–169.
Tomkins AG, Weinberg RF, Schaefer BF, and Langendam A (2013) Disequilibrium melting and melt migration driven by impacts: Implications for rapid planetesimal core formation. *Geochimica et Cosmochimica Acta* 100: 41–59.
Tonks WB and Melosh HJ (1990) The physics of crystal settling and suspension in a turbulent magma ocean. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 151–174. New York: Oxford University Press.
Tonks WB and Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100: 326–346.
Tonks WB and Melosh HJ (1993) Magma ocean formation due to giant impacts. *Journal of Geophysical Research* 98: 5319–5333.
Touboul M, Puchtel IS, and Walker RJ (2012) ^{182}W evidence for long-term preservation of early mantle differentiation products. *Science* 335: 1065–1069.
Turcotte DL and Schubert G (2002) *Geodynamics*, 2nd edn. New York: Cambridge University Press.
Urey HC (1955) The cosmic abundances of potassium, uranium, and thorium and the heat balances of the Earth, the Moon, and Mars. *Proceedings of the National Academy of Sciences of the United States of America* 41: 127–144.
van der Molen I and Paterson MS (1979) Experimental deformation of partially-melted granite. *Contributions to Mineralogy and Petrology* 70: 299–318.
Voorhees PW (1992) Ostwald ripening of two-phase mixtures. *Annual Review of Materials Science* 22: 197–215.
Walker D, Powell MA, Lofgren GE, and Hays JF (1978) Dynamic crystallization of a eucrite basalt. In: *Proceedings of the 9th Lunar and Planetary Science Conference,* pp. 1369–1391.
Walter MJ, Nakamura E, Trønnes RG, and Frost DJ (2004) Experimental constraints on crystallization differentiation in a deep magma ocean. *Geochimica et Cosmochimica Acta* 68: 4267–4284.
Walter MJ and Trønnes RG (2004) Early Earth differentiation. *Earth and Planetary Science Letters* 225: 253–269.
Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240.
Weidenschilling SJ, Spaute D, Davis DR, Marzari F, and Ohtsuki K (1997) Accretional evolution of a planetesimal swarm. 2. The terrestrial zone. *Icarus* 128: 429–455.
Weinstein SA, Yuen DA, and Olson PL (1988) Evolution of crystal-settling in magma-chamber convection. *Earth and Planetary Science Letters* 87: 237–248.
Wetherill GW (1975) Radiometric chronology of the early solar system. *Annual Review of Nuclear Science* 25: 283–328.
Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
Wetherill GW (1990) Formation of the Earth. *Annual Review of Earth and Planetary Sciences* 18: 205–256.
Willis GE and Deardorff JW (1974) A laboratory model of the unstable planetary boundary layer. *Journal of Atmospheric Science* 31: 1297–1307.
Wood JA (1975) Lunar petrogenesis in a well-stirred magma ocean. In: *Proceedings of the 6th Lunar Science Conference,* pp. 1087–1102.
Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites and a geophysical model of the moon. In: Levinson AA (ed.) *Proceedings of the Apollo 11 Lunar Science Conference*, pp. 965–988. New York: Pergamon.
Wright K and Price GD (1993) Computer simulation of defects and diffusion in perovskites. *Journal of Geophysical Research* 98: 22245–22253.
Yamazaki D and Karato S-I (2001) Some mineral physics constraints on the rheology and geothermal structure of Earth's lower mantle. *American Mineralogist* 86: 385–391.
Yamazaki D, Kato T, Yurimoto H, Ohtani E, and Toriumi M (2000) Silicon self-diffusion in $MgSiO_3$ perovskite at 25 GPa. *Physics of the Earth and Planetary Interiors* 119: 299–309.

Yin Q, Jacobsen SB, Yamashita K, Blichert-Toft J, Telouk P, and Albarède F (2002) A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* 418: 949–952.

Yoshino T, Walter MJ, and Katsura T (2003) Core formation in planetesimals triggered by permeable flow. *Nature* 422: 154–157.

Zahnle K, Arndt N, Cockell C, et al. (2007) Emergence of a habitable planet. *Space Science Reviews* 129: 35–78.

Zahnle KJ, Kasting JF, and Pollack JB (1988) Evolution of a steam atmosphere during Earth's accretion. *Icarus* 74: 62–97.

Zaranek SE and Parmentier EM (2004) Convective cooling of an initially stably stratified fluid with temperature-dependent viscosity: Implications for the role of solid-state convection in planetary evolution. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2003JB002462.

Zel'dovich Ya B and Raizer Yu P (2002) *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*. Mineola, NY: Dover Publications.

Zerr A and Boehler R (1993) Melting of $(Mg,Fe)SiO_3$-perovskite to 625 kilobars: Indication of a high melting temperature in the lower mantle. *Science* 262: 553–555.

Zerr A and Boehler R (1994) Constraints on the melting temperature of the lower mantle from high-pressure experiments on MgO and magnesiowustite. *Nature* 371: 506–508.

Zerr A, Diegeler A, and Boehler R (1998) Solidus of Earth's deep mantle. *Science* 281: 243–246.

Zhang J and Herzberg CT (1994) Melting experiments on anhydrous peridotite KLB-1 from 5.0 to 22.5 GPa. *Journal of Geophysical Research* 99: 17729–17742.

9.05 Water in the Evolution of the Earth and Other Terrestrial Planets

S Karato, Yale University, New Haven, CT, USA

© 2015 Elsevier B.V. All rights reserved.

9.05.1	**Introduction**	105
9.05.2	**Equilibrium and Transport Properties (Processes) Relevant to Volatile Distribution**	107
9.05.2.1	General Introduction	107
9.05.2.2	Condensation	108
9.05.2.3	Solubility of Volatiles in the Condensed Matter	110
9.05.2.4	Transport of Volatiles	111
9.05.2.4.1	Diffusion	111
9.05.2.4.2	Melt–solid separation	112
9.05.2.4.3	Macroscopic segregation involving a melt-rich layer in the midmantle	114
9.05.3	**Influence of Water on Physical and Chemical Properties**	114
9.05.3.1	Melting	114
9.05.3.2	Rheological Properties	115
9.05.3.3	Seismic Wave Propagation	117
9.05.3.4	Electrical Conductivity	117
9.05.3.5	Influence of Partial Melting	119
9.05.4	**Distribution of Volatile Elements on Earth**	120
9.05.4.1	Geochemical Inference of Water and Other Volatile Element Distribution on Earth	120
9.05.4.2	Geophysical Inference of Water Distribution	122
9.05.4.2.1	From seismological observations	123
9.05.4.2.2	From electrical conductivity	124
9.05.4.2.3	From rheological properties	125
9.05.4.3	Other Volatile Elements	125
9.05.4.4	Water Content in Other Planetary Bodies	126
9.05.5	**Volatiles During Planetary Formation**	127
9.05.5.1	Acquisition and/or Loss of Volatile Elements in the Early Stage of Planetary Formation	127
9.05.5.1.1	General introduction	127
9.05.5.1.2	The snow line	128
9.05.5.1.3	Water (hydrogen) within the snow line?	128
9.05.5.1.4	Gas → solid versus gas → liquid condensation	128
9.05.5.2	Behavior of Volatiles During the Later Stage of Planetary Formation	129
9.05.5.2.1	Collision-induced melting and vaporization	129
9.05.5.2.2	Formation of the atmosphere and the ocean	131
9.05.5.2.3	Influence of core formation	131
9.05.6	**Evolution of the Ocean and the Atmosphere**	132
9.05.6.1	Observational Constraints	132
9.05.6.2	Stability of the Atmosphere and the Hydrogen (Water) Loss from the Atmosphere	134
9.05.6.3	Evolution of the Ocean Mass: Models of Global Water (Volatile) Circulation	134
9.05.7	**Origin of Volatiles on Earth and Other Terrestrial Planets**	137
9.05.8	**Summary and Outlook**	138
Acknowledgment		139
References		139

9.05.1 Introduction

The Earth is a unique planet. As far as we know, it is the only planet where life has evolved and where plate tectonics operates. It is generally believed that this is in large part due to the presence of a certain amount of liquid water on its surface for most of its history (e.g., Langmuir and Broecker, 2012). Furthermore, the amount of water that Earth has on its surface appears to be just right for life to evolve: if Earth had much more or less water, then life would not have evolved or would have been more difficult to evolve (e.g., Maruyama et al., 2013). Consequently, one of the important questions in the study of the Earth as a terrestrial planet is how did Earth get that amount of water and how has Earth kept the ocean on its surface for such a long time? To what extent is it a natural consequence of formation and evolution of planets, and to what extent did it happen by chance?

To answer these questions, one needs to understand the physics and chemistry of behavior of water and other volatiles during planetary formation and evolution and analyze geologic, geochemical, and geophysical observations to obtain some

observational constraints. The purpose of this chapter is to review these two issues.

Planets are formed as a by-product of star formation (e.g., Hayashi et al., 1985; Safronov, 1972). Therefore, the source materials of planets are essentially the same as that of a star with which they are associated. In the case of the solar system, the composition of the Sun can be estimated from the spectroscopic studies of its outer convective layer and is dominated by the volatile elements such as H, He, C, and N (**Figure 1(a)**; Lauretta, 2011). These are the source materials from which planets were formed. Some of them (H, C, O, N, P, and K) are the key elements of life and others are essential ingredients of rocks (Mg, Si, O, Fe, Ca, and Al). The theory of formation of elements shows that the presence of these elements is a natural consequence of nucleosynthesis during the Big Bang and subsequent star evolution (Burbidge et al., 1957; Langmuir and Broecker, 2012). Indeed, the chemical composition of one of the most primitive materials in the solar system (carbonaceous chondrite) shows a similarity to the solar composition except for the volatile elements that are depleted relative to the solar abundance (**Figure 1(b)**).

Therefore, it is critical to understand how and to what extent volatile elements are depleted during planetary formation and evolution. Recall that the amount of volatile elements to make a 'wet' planet like the Earth is not large. The mass of the ocean on Earth is ~0.023 wt% of the total mass (~0.033 wt% of the mantle). This means that even 99% of depletion of water from the carbonaceous chondrite (likely the source materials for terrestrial planets in the solar system) still makes a planet 'wet' (~5 ocean mass of water). Note that even a small amount (~0.01–0.1 wt%) of volatiles such as hydrogen affects two of the key properties of rocky materials, namely, rheological properties (e.g., Karato and Jung, 2003; Mei and Kohlstedt, 2000a,b) and melting relationships (e.g., Dasgupta and Hirschmann, 2006; Hirschmann, 2006; Inoue, 1994; Kushiro et al., 1968), that control the dynamics and evolution of terrestrial planets. Although a small amount, it is these volatile elements that control the surface conditions of terrestrial planets where life might evolve (e.g., Abe et al., 2011; Langmuir and Broecker, 2012; van Thienen et al., 2007). Consequently, it is critical to understand how much of volatile elements are acquired to a growing planet and how they may be circulated during the long-term evolution of a planet.

Plausible processes that might control the volatile acquisition and circulation in a planet are schematically summarized in **Figure 2** (see also **Table 1**). They include (i) condensation from the primitive solar nebula, (ii) collisions of planetary embryos including later-stage collisions of relatively large planetary bodies ('giant impacts'), (iii) magma ocean formation and crystallization and interaction of magma ocean with the atmosphere, (iv) core formation, and (v) long-term evolution of atmosphere–hydrosphere coupled with mantle convection and core cooling. Each of these processes (or groups of processes) involves microscopic physics and chemistry and meso- to macroscopic processes that control the large-scale distribution of volatiles. **Table 1** provides a brief summary of each process and issues to be discussed in some detail in this article.

Among various volatile elements, hydrogen has strong influence on some of the key physical and chemical properties (e.g., rheological properties of rocks and melting relationships), and consequently, hydrogen distribution has strong feedback to geodynamic processes. The strong influence of hydrogen on

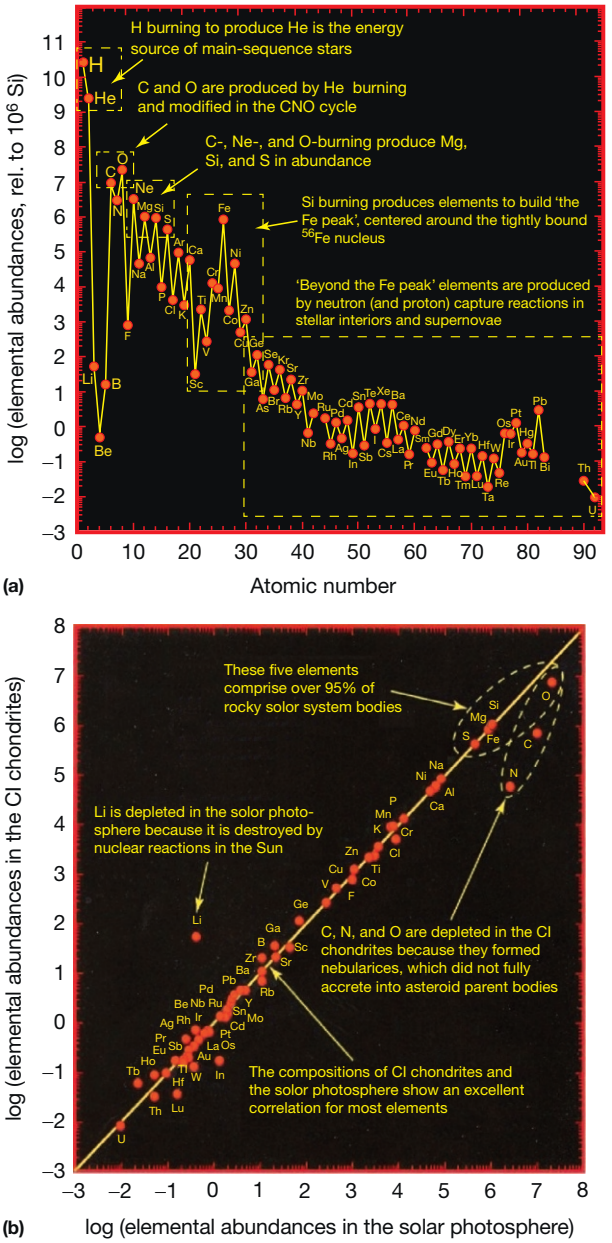

H burning to produce He is the energy source of main-sequence stars
C and O are produced by He burning and modified in the CNO cycle
C-, Ne-, and O-burning produce Mg, Si, and S in abundance
Si burning produces elements to build 'the Fe peak', centered around the tightly bound 56Fe nucleus
'Beyond the Fe peak' elements are produced by neutron (and proton) capture reactions in stellar interiors and supernovae
log (elemental abundances, rel. to 10^6 Si)
Atomic number
(a)
These five elements comprise over 95% of rocky solor system bodies
Li is depleted in the solor photosphere because it is destroyed by nuclear reactions in the Sun
C, N, and O are depleted in the CI chondrites because they formed nebularices, which did not fully accrete into asteroid parent bodies
The compositions of CI chondrites and the solor photosphere show an excellent correlation for most elements
log (elemental abundances in the CI chondrites)
log (elemental abundances in the solar photosphere)
(b)

Figure 1 (a) Cosmic abundance of elements. (b) Abundance of element in carbonaceous chondrite. Cosmic abundance reflects processes of element synthesis, while abundance of elements in carbonaceous chondrite shows systematic depletion in volatile elements (elements with low condensation temperature). Reproduced form Lauretta DS (2011) A cosmochemical view of the solar system. *Elements* 7: 11–16.

some physical properties also allows us to infer hydrogen distribution using some geophysical observations (e.g., Karato, 2011). Therefore, the focus will be placed on hydrogen (water). However, a brief discussion on other volatile elements such as carbon will also be presented where appropriate. The factors controlling volatile distribution include equilibrium thermodynamic properties of volatiles such as the partitioning coefficients of volatile elements between two coexisting phases, diffusion coefficients of volatiles in minerals, and the meso- to macroscopic processes by which materials containing volatile elements are transported.

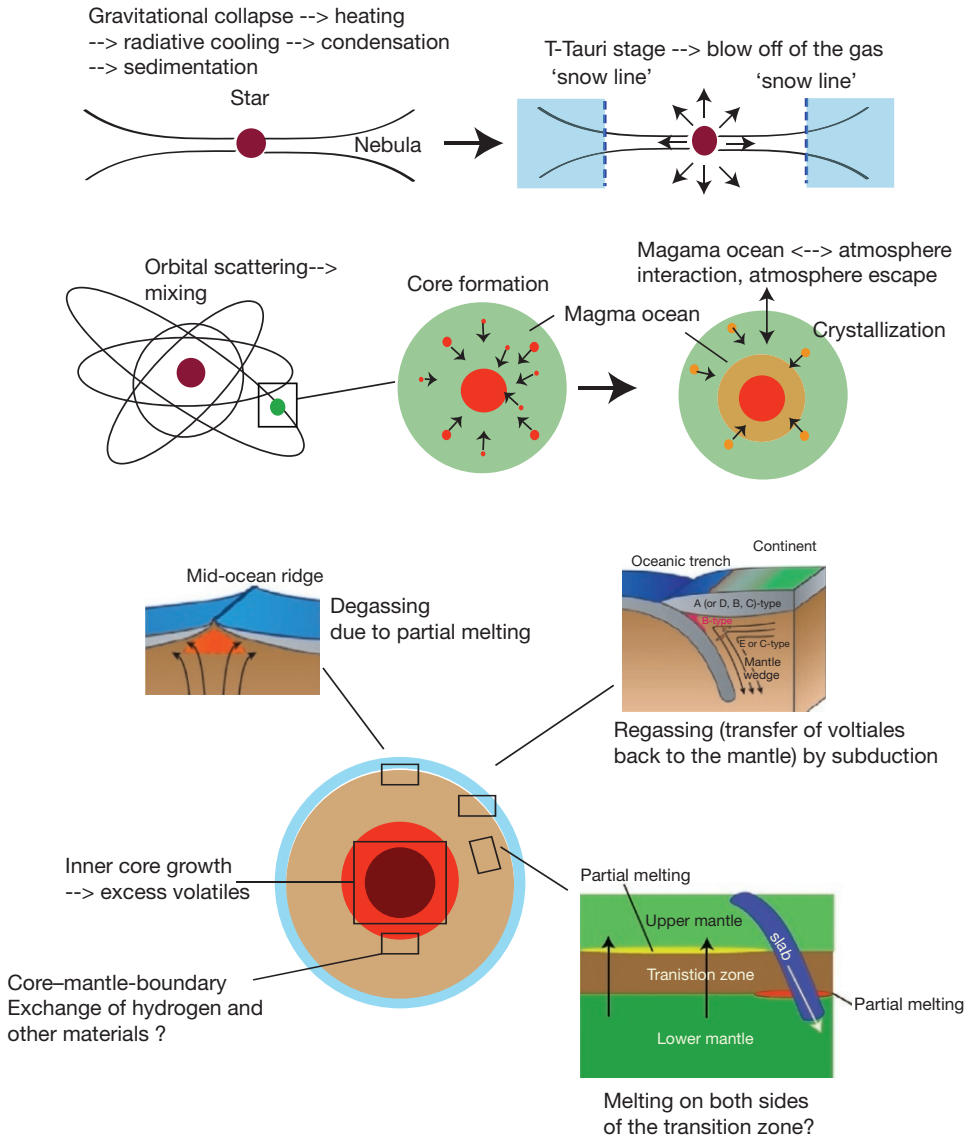


Figure 2 A schematic diagram of possible process controlling the volatile distribution in planets.

There are two aspects in doing such a study. First, it is important to identify the processes by which volatile content (and its distribution) in terrestrial planets might be controlled. Second, it is critical to estimate, based on a range of observations, how much volatiles are in the terrestrial planets and how these volatiles are distributed in space and time (evolution). Because a large amount of data are available for Earth, my discussion will be naturally focused on Earth, but issues of volatiles in other terrestrial planetary bodies including the Moon, Venus, Mars, and meteorites will also be discussed.

9.05.2 Equilibrium and Transport Properties (Processes) Relevant to Volatile Distribution

9.05.2.1 General Introduction

Volatile distribution is affected by phase transformations, notably by those among different states of matter. For a given matter, there are three different states (gas, liquid, and solid), and one can think of transformations among three different combinations of two states: gas↔solid, solid↔liquid, and liquid↔gas (Figure 3). The first one (gas↔solid) is studied in cosmochemistry, while the second one (solid↔liquid) is discussed in geochemistry/petrology. Because the solubility of volatiles is different among different phases (see Section 9.05.2.3), changes in volatile distribution will occur upon these phase transformations. For instance, condensation will reduce the water component unless condensed materials are hydrous minerals such as ice and serpentine. Similarly, by partial melting (solid↔liquid transformation), a substantial amount of volatile is lost from the rock. In contrast to these two processes that have been studied in cosmochemistry and geochemistry (petrology), respectively, the third process, that is, gas↔liquid transformation, has never been studied in detail (except for Ebel and Grossman, 2000). As I will review later (see also Karato, 2013a), the gas↔liquid phase transformation may play an important role in some processes of planetary formation.

Table 1 Various processes in planetary evolution that may control the volatile budget

	Processes	*Questions*
Condensation	Heating by gravitational collapse +radiation from the star Cooling of the protoplanetary nebula Condensation Transport of condensed matter	Where is the 'snow line' (how does it depend on the properties of the star and the nebula)? Do condensed matters ('planetesimals') acquire enough volatiles within the snow line to make a 'wet' planet like the Earth?
Collisions	Degassing by collisions Orbital scattering	How much heating has occurred by collisions? What is the extent of orbital scattering and when?
Magma ocean	Melting Solidification Magma ocean–atmosphere interaction	How does a magma ocean interact with the (proto)atmosphere? To what extent is a planet chemically stratified by the solidification of a magma ocean?
Core formation	Small-scale (equilibrium) separation Large-scale (nonequilibrium) separation	How much volatiles did the core acquire during the core formation? How can we infer the hydrogen content in the core?
Long-term evolution	Mantle convection (phase transformations) Core cooling Degassing (volcanism), regassing (subduction) Atmospheric escape	How are volatile elements distributed in Earth and other terrestrial planets? When was the ocean formed? How has the ocean mass changed with time? How does the regassing (degassing) rate change with time? What is the nature of feedback between mantle convection and mantle (and surface) volatiles? Are there internal processes in the mantle (and core) to regulate the ocean mass? How much is the atmospheric escape (how do we explain D/H of different planetary bodies)?

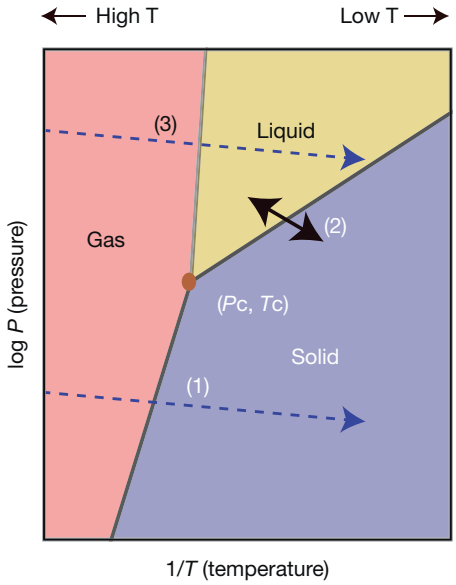


Figure 3 A schematic gas–liquid–solid phase diagram. Gas ↔ solid transitions are usually studied in cosmochemistry. Solid ↔ liquid transitions are usually studied in geochemistry. Gas ↔ liquid transitions are not studied extensively, but will play an important role in a relatively high-pressure environment. Above certain pressure, gas and liquid states become indistinguishable. This condition is referred to as the critical point. This region is not included in this figure because the focus is condensation that occurs at relatively low pressures.

In most cases, the equilibrium properties (e.g., element partitioning coefficients) have important influence on the volatile distribution associated with these processes. Therefore, I will provide a brief review on this subject. However, properties or processes of transport of volatiles must also be considered in order to interpret various observations properly. For example, Keppler and Bolfan-Casanova (2006) and Inoue et al. (2010) discussed that equilibrium partition coefficients of water may control the actual distribution of water, explaining a higher water content in the transition zone than in the upper or the lower mantle by the different hydrogen solubility. Such an argument is unlikely valid, however, because diffusion is very inefficient in transporting elements to a large scale (even hydrogen, the diffusion distance for ~1 Gyr is ~1–10 km in most part of the Earth's mantle). Consequently, without any large-scale mass transport, the distribution of some elements cannot change from one region to another at a scale larger than ~100 km. In other words, if a large difference in composition between different regions (with large size, ~100 km or more) of the Earth (and other terrestrial planets) was identified, it would imply the presence of large-scale processes to segregate materials, for example, partial melting and resultant melt–solid separation.

9.05.2.2 Condensation

At high enough temperatures, all elements assume gaseous state where entropy is large (in various molecular species, H, H_2, CO, SiO, etc.). When a hot gas is cooled, then the energy reduced by forming chemical bonding exceeds the amount of entropy obtained in a gaseous state, and a condensed phase will be formed. Let us consider an element E that condenses from the protoplanetary nebula that is dominated by hydrogen molecules. Let us assume the chemical equilibrium between the gas and a condensed phase containing the element E and also that the element E is a minor element compared with H (hydrogen) and behaves as an ideal gas in the gas phase. Then, applying the law of mass action, one can show that the fraction of condensation of element E, α_E, is related to temperature and total pressure of the gas as (e.g., Grossman and Larimer, 1974; Larimer, 1967)

$$\ln(1-\alpha_E) = -\frac{\Delta G_E}{RT} - \ln f_E - \ln\frac{P_T}{P_o} \quad [1]$$

where f_E is the molar fraction of element E in the gas, ΔG_E is the Gibbs free energy difference of element E between gas state and condensed state (this represents the bonding strength), R is the gas constant, T is temperature, and P_T is the total pressure (P_o is the reference pressure at which ΔG_E is defined). Condensation starts at a certain temperature, T^E_{oo}, where $\frac{\Delta G_E}{RT^E_{oo}} + \ln f_E + \ln\frac{P_T}{P_o} = 0$ and the fraction of condensed matter decreases sharply with further cooling because $\Delta G_E \gg RT$ (Figure 4). In most cases, the *condensation temperature* (T^E_{cond}) of an element E is defined as a temperature at which the fraction of condensed matter is 50% ($\alpha_E = 0.5$):

$$T^E_{cond} = \frac{\Delta G_E}{R\left(0.693 - \ln f_E - \ln\frac{P_T}{P_o}\right)} \quad [2]$$

As one can see from this equation, an element with a small ΔG_E (weak chemical bonding) has a low condensation temperature. These are called volatile elements. Note that the condensation temperature also depends on the total pressure P_T.

Urey was one of the first scientists who studied the physics and chemistry of condensation to understand the chemistry of planets (see Chapter 4 of Urey, 1952). Systematic studies on condensation in the primitive solar nebula were made in the early 1970s (e.g., Grossman, 1972; Grossman and Larimer, 1974; Figure 4). Those calculations were successful in explaining a variety of chemical compositions of meteorites and to some extent the compositions of terrestrial planets (Lewis, 1972, 1974). In particular, Grossman and Larimer (1974) were able to explain a variety of compositions of meteorites including the presence of calcium-rich inclusions in Allende meteorite (carbonaceous chondrite) and the Fe content of some of the meteorites.

However, these calculations are not so successful in explaining the abundance of volatile components. Grossman and Larimer (1974) discussed that the abundance pattern of volatile elements (Ag, Pb, Bi, In, Tl, and Zn) in carbonaceous chondrites cannot be explained by the equilibrium model of condensation by assigning a single condensation temperature and suggested that these meteorites were formed by the mixing of materials with different degrees of volatile depletion (different condensation temperatures). This suggests orbital scattering (variation in the eccentricity of orbits) during the processes of planetary formation as I will discuss later. However, no detailed calculations were made on the condensation of hydrous minerals because of the lack of reliable data on the kinetics of condensation.

There are a few issues on the condensation of volatile elements. The volatility (condensation temperature) of a given element depends on the condensed phase to which the element condenses (for a given element, there are several possible condensed phases). For instance, for hydrogen, if ice (H_2O) is the condensed phase to which hydrogen condenses first, then the condensation temperature for a typical solar nebula conditions is ~170 K (this defines the 'snow line'; see Section 9.05.5). However, hydrogen could condense as other more refractory materials such as serpentine ($Mg_3Si_2O_5(OH)_4$) or brucite ($(Mg(OH)_2$) could. In these cases, the condensation temperature would be much higher, and one would expect to have a nonnegligible amount of water even within the 'snow line.' However, in many of the previous studies, condensation of hydrogen into these hydrous minerals was not included because it was considered that the kinetics to form these phases is slow. Actual condensation temperature of hydrogen is important in understanding how hydrogen is acquired in a growing terrestrial planet. This issue will be discussed later.

By the same token, the volatility (condensation temperature) of S (and Se) depends on the fugacity of hydrogen. Dominant gaseous phase containing S or Se is H_2S or H_2Se, and therefore, at low hydrogen content, the vapor pressure is low that makes S and Se less volatile. Ringwood (1979) explained anomalously high concentration of S and Se in the lunar rocks by a low hydrogen fugacity (compared with the

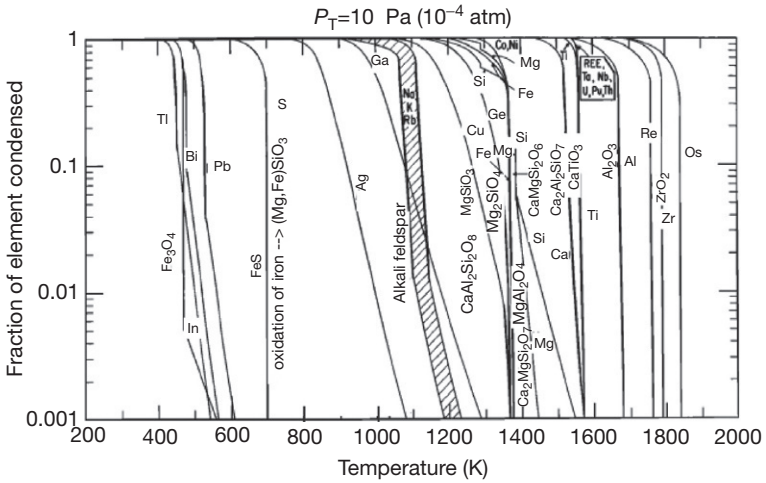


Figure 4 Condensation of various elements. A given element starts to condense at a certain temperature and the fraction of condensed material increases rapidly with decreasing the temperature. 'Condensation temperature' is often defined as a temperature at which 50% of the element is condensed. Reproduced from Grossman L and Larimer JW (1974) Early chemical history of the solar system. *Review of Geophysics and Space Physics* 12: 71–101.

hydrogen fugacity in the primitive solar nebula) in which lunar materials were condensed.

Also important is if one treats the system as an open or a closed system. The question of open versus closed system is partly related to the transport of condensed matter. When condensed materials are formed in the nebula, they will migrate due to gravity and by the pressure caused by radiation. Many of them sink to the equatorial plane but some may migrate to the Sun (e.g., Cassen, 1996). If this migration is effective, then the composition of the system we are considering will change immediately after the condensation of some materials occurs. Orbital scattering changes the eccentricity of orbits of any condensed bodies (particularly the large bodies; Kokubo and Ida, 2000; Morbidelli et al., 2000) and will also affect the composition of planetary materials (see **Section 9.05.5**).

9.05.2.3 Solubility of Volatiles in the Condensed Matter

After a terrestrial planet is formed, the most important processes to redistribute volatile elements are melting and solidification (solid↔liquid transitions) (**Figure 3**). For instance, the magma ocean formed in the early stage of planetary evolution could contain a substantial amount of volatiles. Upon solidification, volatiles will be redistributed, and the way in which volatiles are redistributed is controlled by (i) the partition coefficients (solubility ratios) of volatiles between solids and liquids and by (ii) the mode of solid and liquid separation. Also, after the magma ocean was completely solidified, then partial melting in some regions of the planet will redistribute the volatile elements. Similar to the solidification from the magma ocean, the key factors controlling the volatile distribution upon partial melting are (i) the partition coefficients of volatile elements between solid and liquid and (ii) the mode of liquid–solid separation.

The partition coefficients are the ratio of concentration of an element between two phases and are essentially controlled by the solubility of that element in two phases. Since a gas can dissolve any species to a large degree, the partitioning of an element between the gas phase and a condensed phase (solid or liquid) is essentially controlled by the solubility of that element in the condensed phase. The basic physics of solubility is similar to condensation, but an important difference is that in solubility, we consider an element in a condensed phase as an impurity.

Such a reaction can be written schematically as

$$r{\cdot}Y(\mathrm{X})_{(\mathrm{gas})} + A_{(\mathrm{solid,\ liquid})} \leftrightarrow A(\mathrm{X})_{(\mathrm{solid,\ liquid})} \qquad [3]$$

where Y(X) is a gaseous phase containing a volatile element X (e.g., H_2O), r is a coefficient that depends on the nature of impurity X in the condensed matter and on the dominant volatile species where the element X occurs in the gas, and A(X) is a condensed phase (olivine, basaltic melt, etc.) that contains element X as an impurity. Assuming a small concentration of dissolved element for which the chemical potential of A(X) is given by $\mu_{A(\mathrm{X})} = \mu^0_{A(\mathrm{X})} + RT\log[\mathrm{X}]$, one finds

$$[\mathrm{X}] = B{\cdot}f_{\mathrm{Y}}^{r}(P,T){\cdot}\exp\left(-\frac{\Delta E + P\Delta V}{RT}\right) \qquad [4]$$

where B is a constant, f_Y is the fugacity of the gaseous phase that contains an element X (Y(X)), ΔE (ΔV) is the energy (volume) difference between material A(X) and A, and other symbols have their usual meaning. The exponent r depends on the nature of chemical reaction including the molecular species of the volatile in the gas and the nature of atomic species dissolved in the condensed matter. When hydrogen exists as water in the gas ($Y = H_2O$) and hydrogen is dissolved in the condensed phase as OH, then $r = 1/2$. Both the fugacity of the volatile species ($f_X(P,T)$) and $\exp\left(-\frac{\Delta E + P\Delta V}{RT}\right)$ term strongly depend on pressure and temperature. Consequently, the volatile solubility strongly depends on pressure and temperature.

The solubility of various volatile elements (H, C, etc.) has been determined for both solid minerals and melts. Microscopic mechanisms of dissolution of volatile elements have also been studied using various spectroscopic methods (e.g., infrared spectroscopy) and theoretical calculations, as well as from the parameters in relation [3] (particularly r and ΔV).

The important results are summarized as follows:

(1) Volatile elements such as H and C can be dissolved in condensed phases in a variety of forms (e.g., OH, H_2, H_2O, hydrogen-related point defects, C, CO_2, and carbon-related point defects).
(2) Solubility of volatile elements such as H and C in melts is much higher than that in solids except for hydrous minerals or carbonates (e.g., McMillan, 1994; Ni and Keppler, 2013).
(3) Hydrous minerals and carbonates can contain a large amount of volatiles (to ~10 wt% or more), but they are stable only at relatively low-temperature conditions such as in and/or near subduction zones (**Figure 5**; the actual stability of a given phase depends also on the total chemistry of a rock and could be narrower) (e.g., Ohtani, 2005; Williams and Hemley, 2001).
(4) Under high-temperature and high-pressure conditions, most of H occurs as an impurity (point defects) in 'nominally anhydrous minerals' (minerals that do not have H in their chemical formula, e.g., Mg_2SiO_4) (e.g., Karato, 2008a; Kohlstedt et al., 1996) (**Figure 6**).
(5) The solubility of H in nominally anhydrous minerals is mineral-specific (for review, see Bolfan-Casanova, 2005; Ingrin and Skogby, 2000; Karato, 2008a). For minerals in the Earth's transition zone (~410–660 km depth), the solubility is quite high (to a few wt%). The solubility of H in the upper mantle minerals is up to ~0.1 wt%. The solubility of H in the lower mantle minerals is not precisely known, but for (Mg,Fe)O and (Mg,Fe)SiO_3 perovskite, the solubility of H is comparable with or less than that in the upper mantle minerals. The solubility of H in $CaSiO_3$ perovskite could be higher but not well constrained. As a consequence, there is a marked depth variation in H solubility (**Figure 7**) that promotes partial melting (Bercovici and Karato, 2003; Ohtani et al., 2004). The solubility of H in iron is very large, leading to the molecular ratio of H/Fe ~ 1 (e.g., Fukai and Suzuki, 1986; Okuchi, 1997). The core is potentially the largest hydrogen reservoir on Earth. However, the actual amount of H in the core depends on the processes of core formation, particularly the relative timing of volatile acquisition and core formation (see **Section 9.05.5**; Ohtani et al., 2005; Rubie et al., 2007). Similarly, the actual amount of H in the mantle is likely different from the solubility limit (see **Section 9.05.4**).

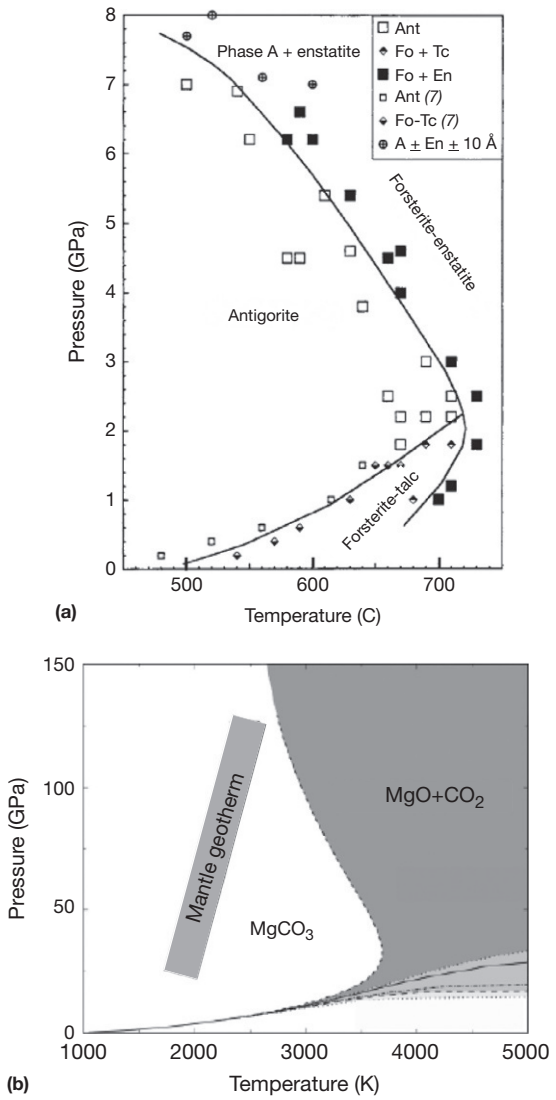


Figure 5 (a) Stability of serpentine (reproduced from Ulmer P and Trommsdorff V (1995) Serpentine stability to mantle depths and subduction-related magmatism. *Science* 268: 858–861). (b) Stability field of magnesite (reproduced from Fiquet G, Guyot F, Kunz M, Matas J, Andrault D, and Hanfland M (2002) Structural refinements of magnesite at very high pressure. *American Mineralogist* 87: 1261–1265).

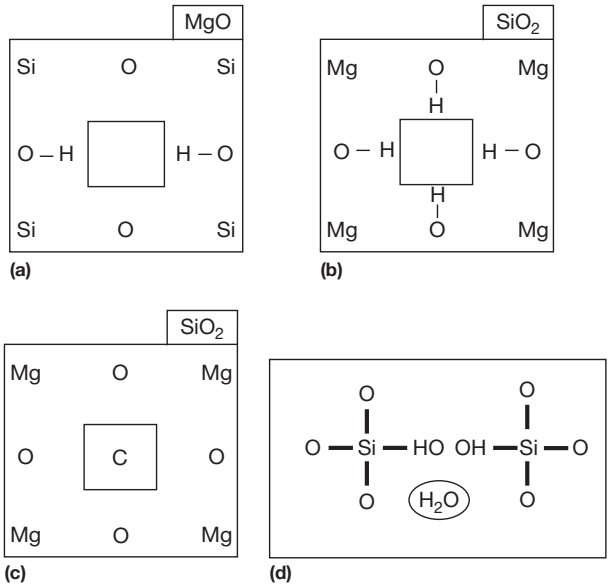


Figure 6 Atomistic models of dissolution of hydrogen and carbon. (a, b) Models of hydrogen dissolution in silicate minerals. (c) A model of carbon dissolution in silicate mineral. (d) Models of dissolution of water in silicate melt (silicate melts could dissolve water either as OH or as H_2O). Reproduced from Karato S (2008a) *Deformation of Earth Materials: Introduction to the Rheology of the Solid Earth*, p. 463. Cambridge: Cambridge University Press.

(6) Hydrogen is dissolved in a mineral in various types of point defects (Nishihara et al., 2008). Hydrogen-related point defects with a small concentration often control the physical properties such as electrical conductivity and rheological properties (Karato, 2006a).

(7) The solubility of carbon in 'nominally ancarbonaceous minerals' (minerals that do not have C in their chemical formula, e.g., Mg_2SiO_4) is much less than that of hydrogen in nominally anhydrous minerals (Shcheka et al., 2006).

(8) The parameters characterizing the solubility (r, ΔE, and ΔV in eqn [3]) depend on the mechanism of dissolution, and consequently, the pressure, temperature, and fugacity dependence of solubility are different among different solubility mechanisms (e.g., Bolfan-Casanova, 2005; Karato, 2008a).

(9) When two (or more) volatile elements (e.g., H and C) coexist, then the solubility of each element into a condensed phase depends on the nature of chemical reaction among them. In this particular case, because both H and C react with O (oxygen), oxygen fugacity also affects the solubility (Ni and Keppler, 2013).

These results form a basis for any discussions on the volatile distribution in terrestrial planets. For instance, the solubility ratios define the element partitioning coefficients that affect how volatiles are redistributed during various processes in planetary evolution. However, these data themselves do not provide direct constraints on the actual distribution of volatile elements. The actual distribution of volatile elements at a large scale (say, >100 km) depends strongly on the processes of evolution of planets because, in many cases, large-scale (>100 km) distribution of elements in a planetary body is not controlled by thermochemical equilibrium. In fact, a numerical study by Richard et al. (2002) showed little or no influence of large water solubility in the transition zone on the nature of global water circulation. This is due to the fact that their model does not include large-scale processes of water redistribution such as partial melting and subsequent melt segregation. In the next section, I will review some properties and processes that may control the actual distribution of volatile elements in a planet.

9.05.2.4 Transport of Volatiles

9.05.2.4.1 Diffusion

Any elements may be transported by diffusion if there is a chemical potential gradient such as the gradient in

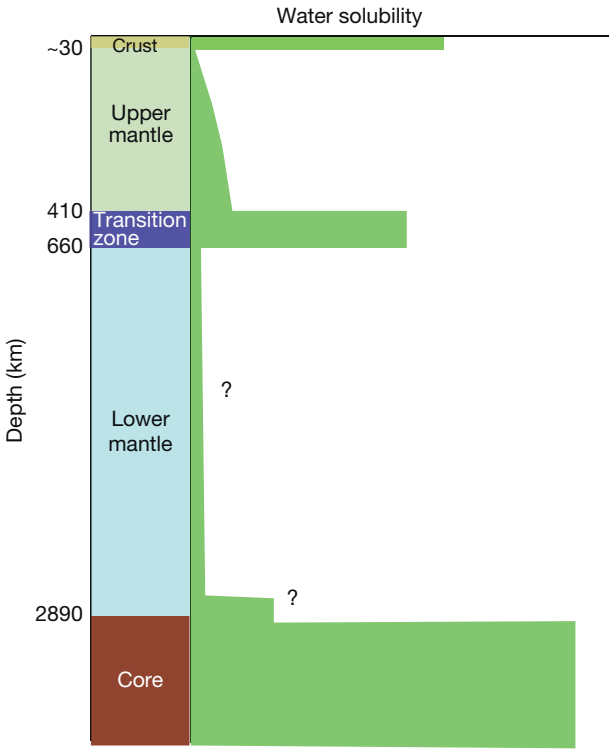


Figure 7 The solubility of water (hydrogen) in the Earth's interior. Water (hydrogen) solubility in minerals changes with depth due to the change in pressure and temperature, chemical layering and to the phase transformations. The water solubility has a peak in the mantle transition zone (~2–3 wt%). The water solubility in the lower mantle minerals is not well constrained. The hydrogen solubility in the core is large. The actual upper limit of hydrogen (water) in minerals can be lower than the solubility if water-bearing phase is not pure water vapor (e.g., a silicate melt). In such a case, the water fugacity in the system is less than the maximum water fugacity of pure water, and the amount of water (hydrogen) that can be present in minerals is smaller. The amount of water (hydrogen) in minerals in these cases is the actual limit of water (hydrogen) content and is often referred to the 'water storage capacity.' The water storage capacity not only is the property of a mineral itself but also depends on the bulk composition of the system.

concentration. Diffusion coefficients of various species in minerals are summarized in Figure 8. Diffusion coefficients vary among various species by as much as ~10 orders of magnitude at the same condition. H (hydrogen) has the fastest diffusion coefficients among them, and hence, in some cases, hydrogen content in minerals corresponds to that at near-equilibrium conditions. However, even for hydrogen, the variation in its concentration at a large scale cannot be controlled by diffusion-controlled chemical equilibrium. To see this, let us calculate the diffusion distance for several atomic species. The length scale of chemical equilibrium attained by diffusion can be given as

$$l \approx 2\sqrt{Dt} \qquad [5]$$

where D is the diffusion coefficient and t is the time. Hydrogen has the highest diffusion coefficients among various elements and hence has the longest diffusion distance.

Figure 9 shows diffusion distance for various elements as a function of timescale. It is seen that in no case diffusion, length becomes larger than ~100 km for the age of the Earth, and it is concluded that chemical heterogeneity with larger than ~100 km space scale will persist and needs to be explained by initial conditions or by transport of matter by processes other than diffusion. However, regarding the chemical equilibrium during partial melting, space scale is different. Here, we consider the exchange of materials at the scale of each grain (several millimeters) when melt–solid separation occurs by percolation. In these cases, chemical equilibrium may be attained for some elements (see Section 9.05.2.4.2).

In defining 'diffusion coefficients,' one should note the difference between isotope (self-) diffusion and chemical diffusion. When hydrogen diffuses from one region of mineral (or rock) to another driven by the concentration gradient, the transport of hydrogen involves the transport of another charged species and it is called chemical diffusion. In such a case, when hydrogen (proton) moves from one region to another, there must be the movement of other species to maintain the charge neutrality. Consequently, the chemical diffusion coefficient is a harmonic average of self-diffusion coefficient of hydrogen and that of another charge-compensating species. In a given mineral, hydrogen can be dissolved in a few different species and each of them has a different diffusion coefficient (mobility). Evidence for multiple hydrogen-bearing species was reported by Nishihara et al. (2008) and the presence of multiple species with different diffusion coefficients explains the observed differences between isotope diffusion of hydrogen (H/D diffusion) and hydrogen-assisted electrical conductivity (see Section 9.05.3.4; Karato, 2013b).

9.05.2.4.2 Melt–solid separation

9.05.2.4.2.1 Compaction

Given the inefficient transport of volatile elements by diffusion even for hydrogen, large-scale transport of volatile elements must involve meso- to macroscopic relative motion of materials. One case is the removal of melt from solid by gravity-induced compaction. When the melt density is different from the solid density, melt tends to be segregated from solid if melt is interconnected. This process is controlled by the melt permeability and plastic deformation of the solid matrix. The processes of melt segregation (compaction) can be characterized by two parameters (e.g., McKenzie, 1984; Ribe, 1985; Richter and McKenzie, 1984): the compaction length,

$$\delta_{\mathrm{comp}} = \sqrt{\frac{k\eta_{\mathrm{s}}}{\eta_{\mathrm{m}}}} \approx 10 \text{ m}, \qquad [6]$$

and the compaction time,

$$\tau_{\mathrm{comp}} = \frac{1}{\Delta\rho \cdot g}\sqrt{\frac{\eta_{\mathrm{m}}\eta_{\mathrm{s}}}{k}} \approx 1 \text{ Myr} \qquad [7]$$

where k is the permeability of melt through solid matrix ($\sim 10^{-17}$ m^2), η_{s} is the viscosity of solid matrix ($\sim 10^{18}$ Pa s), η_{m} is the viscosity of melt (~1 Pa s), $\Delta\rho$ is the density difference between melt and solid (~300 kg m^{-3}), and g is the acceleration due to gravity (~10 m s^{-2}). I conclude that compaction is efficient in most of the cases.

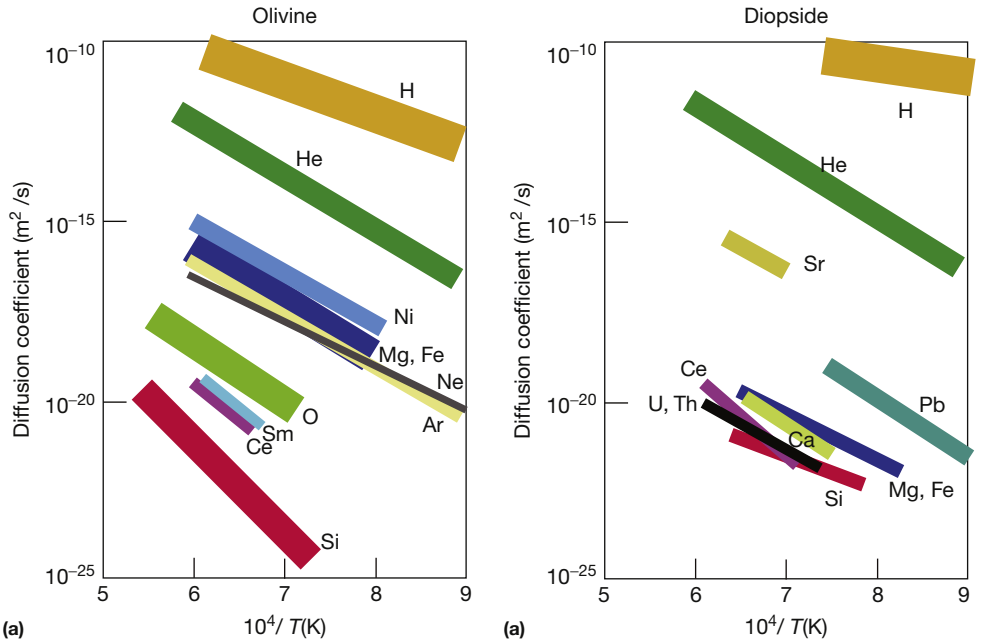


Figure 8 Diffusion coefficients of various elements in olivine ($(Mg,Fe)_2SiO_4$, Fe/(Fe+Mg)$\sim$0.1) and clinopyroxene (diopside ($CaMgSi_2O_6$)) at room pressure or $P\sim1$ GPa. Data: olivine (Baxter, 2010; Chakraborty, 2010; Farver, 2010) and diopside (Baxter, 2010; Cherniak and Dimanov, 2010; Farver, 2010). Results are applicable to shallow regions of the Earth (and other planets, $P<2$–3 GPa). Diffusion coefficients in a mineral depend strongly on diffusing species, and therefore, the degree to which chemical equilibrium is attained during melting (melt segregation) depends on species. Most of 'trace elements' in the upper mantle reside in diopside, and therefore, the diffusion coefficients in diopside control the mode of melt segregation. Hydrogen redistribution upon partial melting is likely controlled by diffusion in olivine (and orthopyroxene).

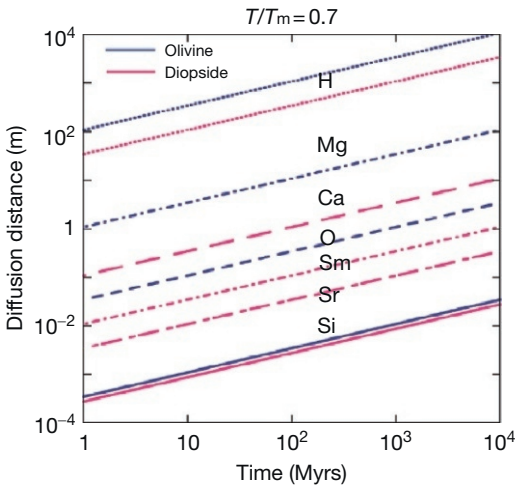


Figure 9 Diffusion distance of various elements in olivine and clinopyroxene (diopside). T, temperature; T_m, melting temperature. $T/T_m=0.7$ is a typical mantle condition. Diffusion distance of hydrogen is plotted as a function of time. In most cases, diffusion distance is small compared to global structures ($\sim$100–1000 km), and any heterogeneity or layering in composition at the global scale implies large-scale material segregation. Diffusion of H at the core–mantle boundary might be an exception. If pressure effects on hydrogen diffusion are small, the diffusion distance of H will be large ($\sim$100 km or more) at the core–mantle boundary and substantial exchange of hydrogen between the core and the mantle.

9.05.2.4.2.2 Chemical equilibrium during melt–solid separation

When melt and solid segregate, chemical equilibrium between them may not be maintained. In order to understand the influence of melt extraction, the mode of melt segregation is often classified into two categories, *batch melting* and *fractional melting*. Batch melting is a process where melt formed from the solid rock migrates slowly in the solid matrix so that whole melt is always in equilibrium with whole solid. So this may be called equilibrium melting. Fractional melting is opposite where melt migrates so fast that chemical equilibrium between melt and solid is attained only locally at the melt–solid interface. Fractional melting can remove incompatible elements (e.g., volatile elements) more efficiently from the original rock. A parameter to characterize the mode of melt segregation is the ratio of diffusion timescale and timescale of melt–solid contact (Iwamori, 1993), that is,

$$\psi \equiv \frac{d^2 \upsilon}{DL} \qquad [8]$$

where d is the grain size, D is the diffusion coefficient of a relevant species, υ is the velocity of solid moving through a partially molten layer, and L is the thickness of the partially molten layer. If $\psi \gg 1$, then diffusion is so slow that melt will migrate through solid without much exchange of materials (fractional melting). If, on the other hand, $\psi \ll 1$, then chemical equilibrium will be attained during the melt separation (batch melting) (Figure 10). It can be seen that the mode of melt segregation strongly depends on diffusion coefficients.

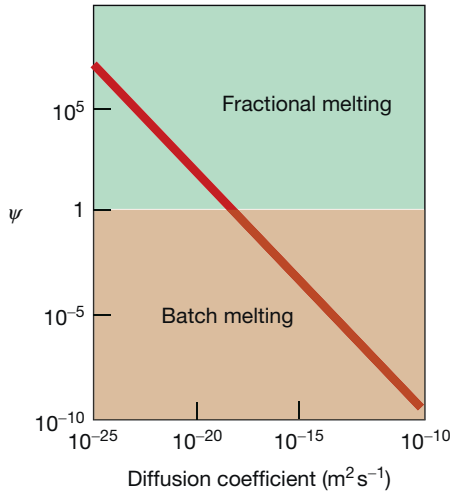


Figure 10 Mode of melt–solid separation. The mode of melting can be classified into batch (equilibrium) melting and fractional (disequilibrium) melting. Whether batch melting or fractional melting occurs is largely controlled by a nondimensional parameter ψ that is a ratio of timescale of diffusion and melt migration through a solid column. If $\psi \ll 1$, then the batch melting occurs; otherwise, fractional melting occurs. Because this parameter depends strongly on the diffusion coefficients and diffusion coefficients vary largely among different species, the mode of melting is likely different among different atomic species. Grain size is 5 mm, velocity of solid material is 5 cm year^{-1}, and the thickness of partially molten layer is 50 km.

An important conclusion is that either batch (equilibrium) or fractional (disequilibrium) melting depends on the atomic species. This figure (also **Figure 9**) shows that for some slow-diffusing species, the melt produced by partial melting may not be as much enriched with these elements even though they are incompatible. These include U, Th, and Ce.

A similar question can be asked as to the mode of solid–liquid separation during freezing of a melt layer (e.g., magma ocean) (or during the core formation (separation of iron from silicates); Karato and Murthy, 1997; Stevenson, 1990). As far as the density of solid is different from that of the melt, separation tends to occur. However, convection current tends to suspend grains if the grain size is small. Such an issue was analyzed by Solomatov (2009) and Tonks and Melosh (1990). If melt–solid separation occurs efficiently during the freezing stage of a magma ocean, then one would expect a substantial layering in composition including the water content (e.g., Elkins-Tanton and Grove, 2011). However, the predicted layering in density for the Moon by Elkins-Tanton and Grove (2011) is not consistent with the observed moment inertia (Konopliv et al., 2001) that indicates that the lunar mantle has nearly homogeneous density.

9.05.2.4.3 Macroscopic segregation involving a melt-rich layer in the midmantle

Somewhat different processes will work if a melt-rich layer is formed inside of the mantle. Such a layer is likely on top of the 410 km and below the 660 km discontinuities in the Earth's interior (e.g., Bercovici and Karato, 2003; Ohtani et al., 2004; Young et al., 1993). In these cases, a thin melt-rich layer (thickness ~ compaction length) will be formed, and the melt will either remain there or be entrained by subducted slabs and carried into the deep mantle (Leahy and Bercovici, 2007). The strongest vertical flow is a downgoing flow associated with subducted slabs, and consequently, these melts or melt-rich materials could go into the deep mantle. The consequence of such a material segregation on the global volatile circulation and hence the evolution of ocean mass will be discussed in **Section 9.05.5**.

9.05.3 Influence of Water on Physical and Chemical Properties

9.05.3.1 Melting

The role of volatiles to enhance melting and to change the composition of melt was already recognized by Bowen (1928). The essential reason for a large influence of volatiles on melting behavior is the large contrast in the solubility of volatile components into liquid and solid. Volatile elements are much more soluble in liquid than solid, and this increases the configurational entropy of liquid thereby reducing its free energy. Consequently, liquid becomes stable relative to solid at high temperatures, enhancing the melting. Assuming the ideal solution model for simplicity, such a behavior can be summarized by the following relationship (e.g., Carmichael et al., 1974):

$$T_m = \frac{T_{mo}}{1 + \frac{RT_{mo}}{\Delta H}\left|\log\frac{1-x_m}{1-x_s}\right|} \quad [9]$$

where T_{mo} is the melting temperature of a volatile-free system, T_m is the melting temperature with x_m being the mole fraction of volatiles in the melt, x_s is the solid, and ΔH is the enthalpy difference between the melt and solid (for a pure system). Because the main cause of the melting temperature depression is the configurational entropy (expressed by the $R\log(1-x)$ term), any impurities would have similar effects as far as their number density is the same.

Extensive experimental studies have been made on the role of volatiles on melting including the role of hydrogen (water) (e.g., Inoue, 1994; Kushiro et al., 1968; Litasov and Ohtani, 2007; **Figure 11**), carbon (e.g., Dasgupta and Hirschmann, 2006; Wyllie and Huang, 1976), and potassium (K) (Wang and Takahashi, 2000).

An important point about melting is that because the melting is enhanced by the presence of volatiles due to the solubility contrast of volatiles between melt and solid, melting caused by volatiles will occur only when the volatile content exceeds a critical value that is proportional to the solubility of volatiles in solids. This means that when there are some solids that can host a large amount of volatiles, melting is difficult. For carbon, its solubility in minerals depends on oxygen fugacity, and hence, a change in oxygen fugacity can change the melting temperature (Rohrbach and Schmidt, 2011). Similarly, if some K-bearing minerals are stable, then K will not enhance melting (Wang and Takahashi, 2000).

After partial melting, one obtains solid and melt whose compositions are different from the composition of the original solid. **Figure 12** is a schematic diagram showing the evolutional trend of composition of coexisting melt and solid. When a material ascends, decreased pressure promotes melting.

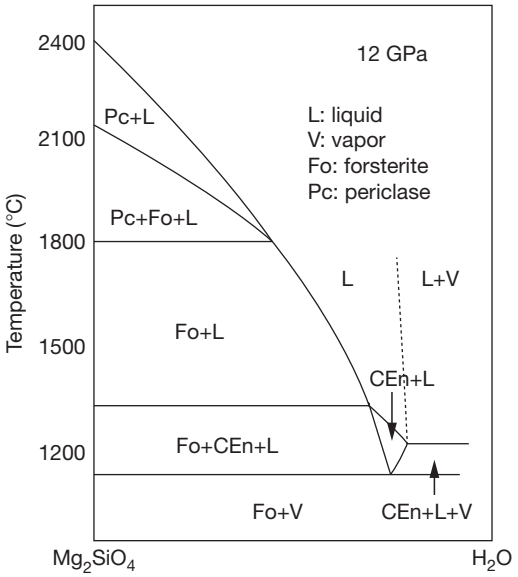


Figure 11 The influence of water on the melting relationship in Mg_2SiO_4. Reproduced from Inoue T (1994) Effect of water on melting phase relations and melt composition in the system Mg_2SiO_4–$MgSiO_3$–H_2O up to 15 GPa. *Physics of Earth and Planetary Interiors* 85: 237–263.

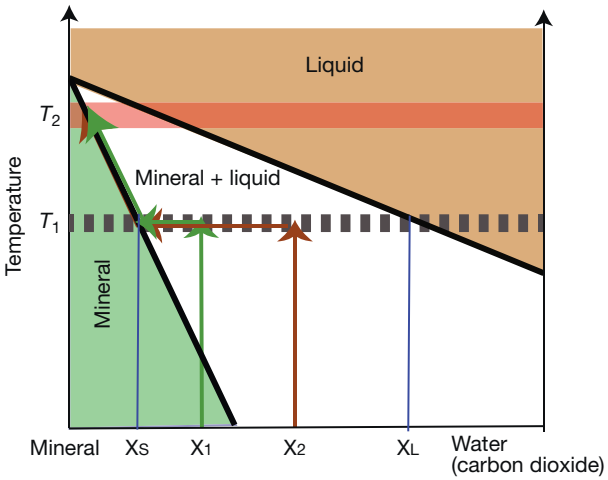


Figure 12 A schematic phase diagram showing volatile-assisted melting. Above the eutectic temperature (T_e), partial melting will occur, if the volatile concentration exceeds the volatile concentration at the solidus. The volatile concentration at the solidus depends on the volatile solubility in the solid. The amount of melt produced by the help of a volatile is determined by the volatile concentration. When a material will move upward below near the surface, the solidus decreases.

When melting involves volatile elements (such as H and C), they will be redistributed because their solubility in melt and solid is different: volatile elements in the solid will be reduced after the melting. However, the volatile content in the solid after the melting will not depend on the initial volatile content so much. Initial volatile content affects the degree of melting but the composition of the residual solid is controlled by the composition of the solidus. Consequently, partial melting homogenizes the composition of the residual solids.

9.05.3.2 Rheological Properties

Griggs and his coworkers were the first to report the strong influence of water (hydrogen) on plastic deformation of minerals. They discovered that the dissolution of hydrogen reduces the creep strength of quartz dramatically (Griggs and Blacic, 1965). A similar phenomenon, although to a lesser degree, was soon observed for olivine (Blacic, 1972). The fundamental reason for the weakening is that the hydrogen dissolved in these minerals makes chemical bonding weak, thereby promoting plastic deformation that requires large-scale relative motion of atoms in a crystal.

Systematic studies on the influence of water on plastic deformation were conducted on several minerals, and the flow law of the following form is usually used to describe the influence of water on plastic deformation (e.g., Karato, 2008a; Kohlstedt, 2009):

$$\dot{\epsilon} = \dot{\epsilon}_{d0} \cdot \sigma^{n_d} \cdot \exp\left(-\frac{H_{cd}^*}{RT}\right) + \dot{\epsilon}_{w0}\left(\frac{C_W}{C_{W0}}\right)^{r_c} \cdot \sigma^{n_w} \cdot \exp\left(-\frac{H_{cw}^*}{RT}\right) \quad [10]$$

where $\dot{\epsilon}$ is the strain rate, $\dot{\epsilon}_{d0(w0)}$ are the preexponential terms for dry (wet) conditions, σ is the stress, $n_{d(w)}$ is the stress exponent, and $H^*_{cd(cw)}$ is the activation enthalpy for dry (wet) conditions. The influence of water to enhance deformation has been demonstrated for all crustal and upper mantle minerals (quartz (Post et al., 1996), olivine (Karato and Jung, 2003), (Mei and Kohlstedt, 2000a), orthopyroxene (Ross and Nielsen, 1978), clinopyroxene (Chen et al., 2006), pyrope garnet (Katayama and Karato, 2008a), and wadsleyite (Kawazoe et al., 2013)). In general, the degree of water weakening is larger for minerals with higher SiO_2 content.

The most detailed studies have been made on olivine, for which water-weakening effect was found for diffusion creep (Karato et al., 1986; Mei and Kohlstedt, 2000a), power-law dislocation creep (high-temperature dislocation creep) (Karato and Jung, 2003; Mei and Kohlstedt, 2000b), and the Peierls mechanism (low-temperature dislocation creep) (Katayama and Karato, 2008b) (**Figure 13**). Water reduces the effective viscosity of olivine up to ~3–4 orders of magnitude for a given stress.

However, water-weakening effect on lower mantle minerals has not been demonstrated. The influence of hydrogen on the diffusion of Mg–Fe in (Mg,Fe)O was studied, but its effect is small at low pressures ($P < 20$ GPa) (Otsuka and Karato, 2013). In lower mantle minerals such as (Mg,Fe)O, the concentration of ferric iron is large and it is unclear if hydrogen has large effects on defect-related properties including plastic deformation and electrical conductivity (Otsuka et al., 2013).

The mechanisms of weakening effects by water (hydrogen) were reviewed by Karato (2008a) and Kohlstedt (2009). Since high-temperature creep is controlled by diffusion-related processes, enhancement of creep by water is partly due to enhanced diffusion (e.g., Costa and Chakraborty, 2008). However, enhancement of diffusion itself cannot explain the strong anisotropy in the water effects on dislocation creep in olivine (Karato, 2010a). A substantial part of water (hydrogen) weakening must be through the modifications to jog/kink densities

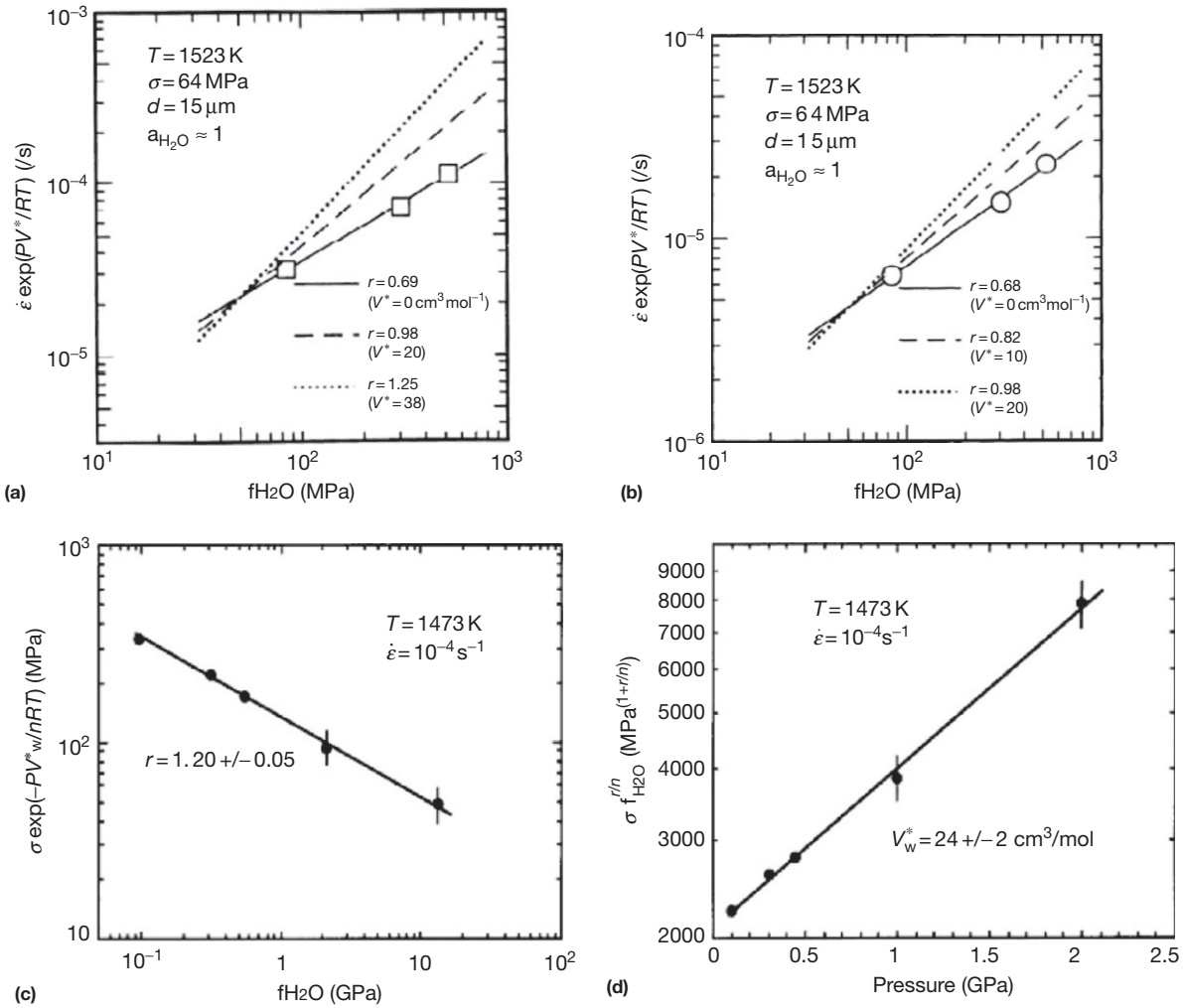


Figure 13 Influence of water on the creep strength of olivine in the (a) dislocation creep regime and (b) diffusion creep regime (after Karato and Jung, 2003; Mei and Kohlstedt, 2000a,b). In the Mei and Kohlstedt (2000a,b) study, experiments were made under water-saturated conditions at low pressures ($P < 0.45$ GPa). Enhancement of deformation with increase of pressure was observed and it was interpreted as a result of increased water fugacity. However, the influence of water fugacity and the intrinsic effect of pressure were not separated, and the key parameters characterizing the water effects were unconstrained (note that the activation volume, V^*, that characterizes the pressure effect remained an arbitrary parameter in their paper). Karato and Jung (2003) conducted deformation experiments at higher pressures (to 2 GPa) and were able to determine the two key parameters (water fugacity exponent (r) and the activation volume (V^*)) for dislocation creep separately. It is necessary to constrain both of these parameters in order to extrapolate these results to the deep upper mantle (deeper than ~30 km) where the influence of water is potentially important (see also Karato, 2010b).

on dislocations. This means that the degree of enhancement of dislocation creep by water is likely larger than the degree of enhancement of diffusion creep.

Influence of other volatile elements such as carbon on plastic deformation is unknown. In olivine, carbon is dissolved in the Si-site (Keppler et al., 2003) and its solubility is low compared with that of hydrogen. However, since Si diffusion likely affects the rheological properties and the concentration of intrinsic defects at Si-site is very small, even a small amount of carbon might have some effects on rheological properties.

The influence of water (hydrogen) to enhance plastic deformation is sometimes anisotropic: hydrogen enhances deformation along one orientation more than along other orientations. Anisotropic water effect can be understood if the influence of water on dislocation properties such as kink/jog density plays an important role (diffusion in olivine is nearly isotropic). This leads to the variation in the lattice-preferred orientation (LPO) (nonrandom distribution of crystallographic orientations of minerals) when plastic deformation occurs by the glide of crystal dislocations (Karato et al., 2008). Consequently, water affects seismic anisotropy in the upper mantle.

Recently, Fei et al. (2013) reported relatively small effects of water on diffusion of Si in Mg_2SiO_4 olivine and argued that creep in olivine is not as much sensitive to water as previously considered. However, such a conclusion is inconsistent with the well-documented large effects of water to enhance creep in olivine ($(Mg,Fe)_2SiO_4$) including anisotropic effects summarized in the preceding text. A more reasonable conclusion from such an observation is either (i) the influence of water is not only through diffusion but also through jog formation

(which results in anisotropic effects) or (ii) the influence of water is different between Mg_2SiO_4 and $(Mg,Fe)_2SiO_4$ olivine.

9.05.3.3 Seismic Wave Propagation

Seismology provides the most detailed information of the Earth's interior (also lunar interior). Water and other volatile elements may affect the seismic wave propagation. If such effects were demonstrated, then water distribution on Earth (and other planets) would be inferred from seismological observations. Based on these ideas, a number of studies have been carried out to investigate how water affects seismic wave propagation. The following aspects of seismic wave propagation have been studied: (i) (average) seismic wave velocities, (ii) topography of velocity discontinuities, (iii) seismic anisotropy, and (iv) seismic wave attenuation.

Dissolution of water (hydrogen) weakens chemical bonding, and hence, one may expect that seismic wave velocities will be reduced by the dissolution of water (hydrogen). However, experimental studies showed that the effects of water on (high-frequency) elastic properties are small (e.g., Jacobsen, 2006; Jacobsen et al., 2008; Mao et al., 2008). In fact, the influence of other factors such as a variation in major element chemistry (e.g., pyrolite versus harzburgite) is much larger than the influence of hydrogen, making it difficult to resolve hydrogen content from seismic wave velocities (Karato, 2011) (**Figure 14**).

High water solubility in transition zone minerals means that the dissolution of water (hydrogen) stabilizes the transition zone minerals relative to the upper or lower mantle minerals. Consequently, the depth of '410 km' or '660 km' discontinuity will be modified by the addition of water. Experimental studies confirmed this (e.g., Chen et al., 2002; Litasov and Ohtani, 2007; Smyth and Frost, 2002), but the magnitude of topography is reduced at high temperatures, and under most of mantle conditions, such topography will be small and hard to detect seismologically (Frost and Dolejš, 2007) (**Figure 15**). In these cases, the actual topography is not much caused by the variation in water content but by the variation in temperature and major element chemistry. Consequently, the inference of water content from these seismological observations is difficult, and sometimes, one finds puzzling results when such a method is used to infer water content (see **Section 9.05.4**).

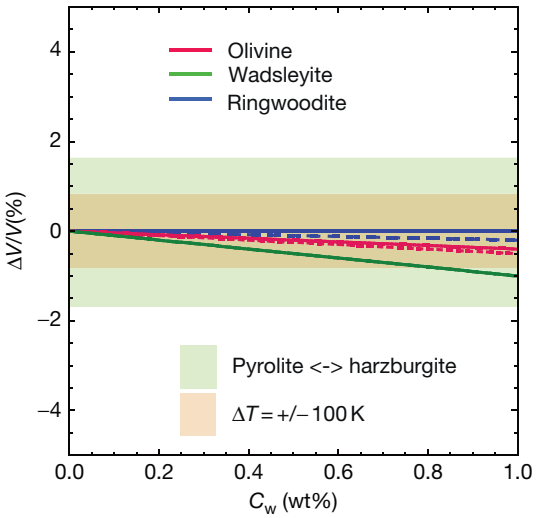


Figure 14 Influence of water on seismic wave velocities at high frequencies. C_w is the water content. Addition of water (hydrogen) reduces seismic wave velocities. However, for plausible water content in the Earth's mantle (<0.1 wt%), its influence is small compared with that of major element chemistry and temperature. The influence of water will be larger at seismic frequencies because of the enhancement of anelastic relaxation.

In contrast, the influence of water on seismic anisotropy in the upper mantle is strong: in some cases, the direction of anisotropy changes substantially (90°). This is caused by the change in the dominant slip system(s) in olivine, and consequently, the distribution of seismic anisotropy can be used to obtain some insight into the distribution of water (Karato et al., 2008) (**Figure 16**).

Also, the addition of water (hydrogen) increases attenuation through the enhanced plastic deformation (decreases Q). This was proposed by Karato (1995) using an analogy with plastic deformation and later confirmed qualitatively by Aizawa et al. (2008) (**Figure 17**). Using essentially the same line of argument, McCarthy et al. (2011) suggested that seismic wave attenuation increases with water content. The functional relationship between seismic wave attenuation and the water content has not been well characterized. However, using the Maxwell time scaling, the following relationship is expected (e.g., Karato, 2006b; McCarthy et al., 2011):

$$Q^{-1} = Q_d^{-1}\exp\left(-\frac{H_{Qd}^*}{RT}\right) + Q_w^{-1}\left(\frac{C_W}{C_{Wo}}\right)^{r_Q}\exp\left(-\frac{H_{Qw}^*}{RT}\right) \quad [11]$$

where Q^{-1} is the fraction of energy loss, $Q_{d(w)}^{-1}$ are the reference values of Q^{-1} for dry (wet) conditions, $H_{Qd(Qw)}^*$ are the activation enthalpy for dry (wet) conditions, and r_Q is a non-dimensional parameter (between 1 and 2).

Shito et al. (2006) developed a method to infer the water content (and temperature and major element chemistry) from the observed velocity and attenuation anomalies. They reported that the deep upper mantle in the Philippine Sea upper mantle has large water content. However, the challenge in using these observations is that the resolution of seismological observations (anisotropy and attenuation) is not as good as the average velocity or topography of discontinuities.

9.05.3.4 Electrical Conductivity

Most minerals are ionic crystals, and therefore, there are always a large number of charged species. However, in an ideal state, most charged species (such as Mg^{2+} and O^{2-}) are strongly bonded to each other, and they do not carry much current except at very high temperatures (close to melting temperature). Impurities such as ferric iron and proton substantially enhance electrical conductivity because electrons attached to ferric iron and protons are both highly mobile (for a review, see Karato and Wang, 2013).

When the concentration of these impurities ('point defects') is increased, electrical conductivity can increase (or decrease) by orders of magnitude. Therefore, electrical conductivity is a good sensor for the physical/chemical state of a planet related to these defects including hydrogen. In many cases, geologic

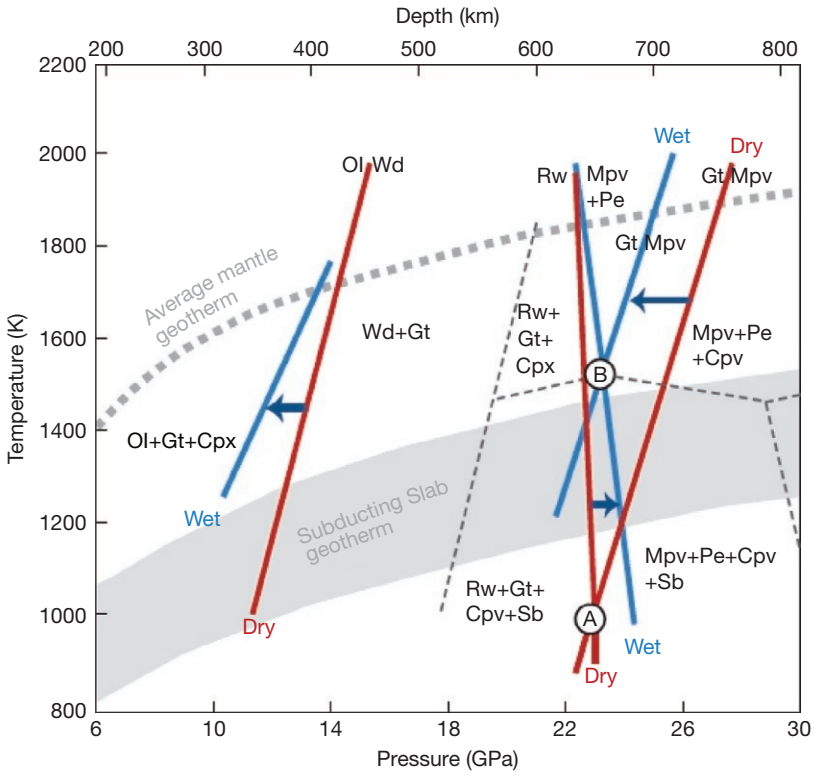


Figure 15 The influence of water on the phase relationship in $(Mg,Fe)_2SiO_4$. Reproduced from Litasov KD and Ohtani E (2007) Effect of water on the phase relations in Earth's mantle and deep water cycle. In: Ohtani E (ed.) *Advances in High-Pressure Mineralogy*, pp. 115–156. Geological Society of America.

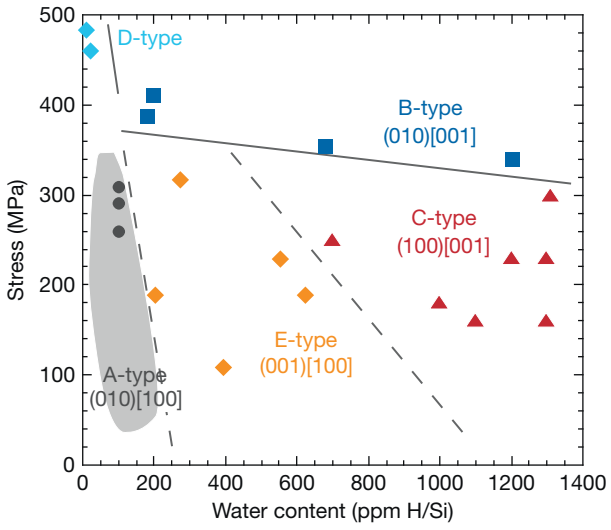


Figure 16 The influence of water (and stress) on the lattice-preferred orientation (LPO) of olivine (after Karato et al., 2008). A, B, C, D, and E types of LPO represent different types of olivine LPO. The LPO of olivine depends on various physical and chemical parameters including water content (water fugacity), stress, temperature, and pressure. The results shown correspond to the LPO at $P < 3$ GPa. At higher pressures, different LPOs might develop. The water content shown is estimated using the Paterson calibration. If SIMS calibration is used, these values should be multiplied by a factor of ~3.

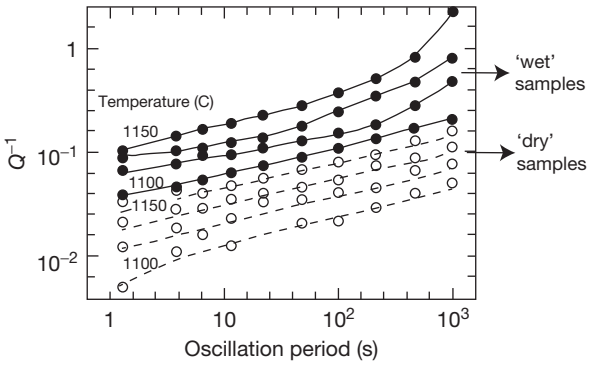


Figure 17 The influence of water on seismic wave attenuation in olivine (Aizawa et al., 2008). Q^{-1} is a measure of seismic wave attenuation (fraction of energy dissipated as heat). 'Wet' samples are those that contain more water than 'dry' due to the use of different jackets. Therefore, the results shown here strongly suggest that water enhances seismic wave attenuation. However, the functional relationship between water content and seismic wave attenuation has not been quantified experimentally.

question we would like to address is the order of magnitude of concentrations of these defects (e.g., hydrogen) as opposed to the precise value of defect concentration. Consequently, although the electrical conductivity of the Earth's interior (or other planetary interiors) cannot be determined as precisely as

seismic wave velocities, even a crude estimate of electrical conductivity provides an important constraint on hydrogen content.

In most minerals, electrical conductivity is caused by two types of impurities, ferric iron and proton (hydrogen). In both cases, the concentration and mobility of charged species are highly dependent on temperature and water content (e.g., Karato, 2008a). Consequently, electrical conductivity of a mineral (σ; not stress!) may be written as

$$\sigma = \sigma_{d0}\exp\left(-\frac{H_{\sigma d}^{*}}{RT}\right) + \sigma_{w0}\left(\frac{C_W}{C_{W0}}\right)^{r_\sigma}\exp\left(-\frac{H_{\sigma w}^{*}}{RT}\right) \quad [12]$$

where $\sigma_{d0(w0)}$ is the preexponential term for dry (wet) mechanism, $H^{*}_{\sigma d(\sigma w)}$ is the activation enthalpy for dry (wet) mechanism, and r_σ is a constant (0.5–1). In most cases, electrical conductivity is also dependent on oxygen fugacity, and in these cases, the influence of oxygen fugacity can be included in the preexponential factor, $\sigma_{d0(w0)}$. Major element chemistry also affects electrical conductivity. Particularly important is the influence of Fe content (in most minerals; Cemič et al., 1980) and Al_2O_3 content in orthopyroxene (Huebner et al., 1979).

Parameters in relation [12] have been determined for some mantle and crustal minerals including plagioclase (Yang et al., 2012), olivine (Wang et al., 2006; Yang, 2012), orthopyroxene (Dai and Karato, 2009a; Yang et al., 2012), clinopyroxene (Yang, 2012), pyrope (Dai and Karato, 2009b), wadsleyite (Dai and Karato, 2009c), and ringwoodite (Huang et al., 2005). Also, the dependence of these parameters on major element chemistry has been determined for some minerals.

The magnitude of the effect of hydrogen is large in comparison with the influence of other factors such as temperature, major element chemistry, and oxygen fugacity (Karato, 2011) (**Figure 18**). Detailed studies on olivine showed that the enhancement of electrical conductivity by the dissolution of hydrogen is not equally caused by all the hydrogen atoms dissolved in a mineral but rather mostly due to a small fraction of dissolved hydrogen with high mobility and that the hydrogen-related species controlling the conductivity changes with temperature (e.g., Dai and Karato, 2014; Karato, 2006a, 2013b; Wang et al., 2006).

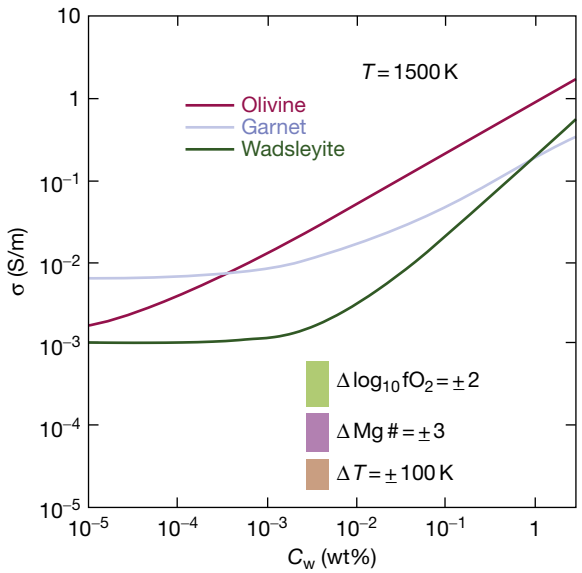


Figure 18 Influence of water on the electrical conductivity of minerals compared with the influence of major element chemistry, temperature, and oxygen fugacity (after Karato, 2011). C_w is the water content. The effect of water (hydrogen) is much larger than the effects of major element chemistry, temperature, and oxygen fugacity for a plausible range of these variables.

The influence of hydrogen on the electrical conductivity has not been studied for the lower mantle minerals. Although there are several studies on the electrical conductivity in lower mantle minerals (e.g., Dobson and Brodholt, 2000; Katsura et al., 1998; Peyronneau and Poirier, 1989; Shankland et al., 1993), the hydrogen content in these samples was not measured, and we do not know if hydrogen has any effect on the electrical conductivity in these minerals (hydrogen was not added in these studies, but some hydrogen is likely dissolved in these minerals). In lower mantle minerals, other defects such as ferric iron and oxygen vacancies contribute to electrical conductivity (e.g., Dobson, 2003), and it is not clear if hydrogen enhances electrical conductivity. It is also noted that most of hydrous minerals do not have high electrical conductivity that is likely due to the low mobility of hydrogen in these minerals (Karato and Wang, 2013).

The influence of carbon on electrical conductivity was discussed by Duba and Shankland (1982). A detailed experimental study was performed by Wang et al. (2013) who showed that when more than ~1 wt% of graphite is present in an olivine aggregate, electrical conductivity is enhanced. The reported high-conductivity regions in the shallow continental lithosphere (e.g., Jones et al., 2003) might represent regions of anomalously high carbon content with low oxygen fugacity.

9.05.3.5 Influence of Partial Melting

Partial melting is ubiquitous in the Earth's mantle when volatile components are present (e.g., Hirschmann, 2010; Kushiro, 2001). Partial melting has effects on various physical properties similar to hydrogen: partial melting makes materials softer (except for the effects of hydrogen removal from minerals that will make materials stronger; see Karato, 1986). Therefore, it is critical to distinguish the influence of volatile element (such as hydrogen) and that of partial melting.

One way to assess the possible contribution from partial melting is to estimate how much melt one needs to explain geophysical anomalies and to address if it is feasible to maintain a required amount of melt. For example, for a given seismic wave velocity reduction, one can estimate the necessary volume fraction of melt that depends on the melt geometry (wetting (dihedral) angle). For the typical asthenosphere (~50–100 km depth in Earth), the dihedral angle is 20–40° (Cooper and Kohlstedt, 1982; Yoshino et al., 2009). For this wetting geometry, the inferred velocity reduction of 5–10% at the lithosphere–asthenosphere boundary requires 3–6% of melt (Takei, 2002). Such a melt fraction is difficult to maintain in a large region from both the static (equilibrium thermodynamics) point of view and dynamic point of view (melt generation and transport) (Karato, 2014b). The essential reasons are (i) because partial melting in many cases in the Earth's mantle is caused by volatiles, the degree of melting is controlled by the amount of volatiles and is small (<0.1%), and

(ii) high melt fraction is difficult to maintain because of the effective compaction (see Section 9.05.2.4.2). An exception is a region where melt completely wets the grain boundaries. In these cases, grain-boundary relaxation would occur that reduces seismic wave velocities by several % (e.g., Karato, 2012). This is a possible case for the deep upper mantle above 410 km where continuous melt production by the upwelling of water-rich materials from the transition zone combined with complete wetting will produce a thick low-velocity region (Karato, 2012; Yoshino et al., 2007).

For electrical conductivity, a smaller melt fraction could explain observed high conductivity if carbonatite melt with a large concentration of alkali ions (Na^+ and K^+) is present (e.g., Gaillard et al., 2008; Naif et al., 2013; Sifré et al., 2014). However, both the magnitude and anisotropy observed by Baba et al. (2006) and Naif et al. (2013) are in good agreement with a hybrid model of hydrogen-assisted conductivity where it is predicted that the conduction mechanism changes at a temperature of ~1300 K to conduction by two protons trapped at M-site and that high anisotropy is consistent with the observed anisotropy (Dai and Karato, 2014; Karato, 2013b). In contrast, the amount of melt needed to explain the observed conductivity exceeds substantially the degree of melting caused by the volatile components (see, e.g., Ni et al., 2011).

It is also noted that a small fraction of melt does not affect the rheological properties strongly under the shallow mantle conditions (Kohlstedt, 2002). The small effect of partial melt is attributed to the wetting geometry of basaltic melt under the shallow upper mantle conditions (Kohlstedt, 2002) and is similar to the small effect on seismic wave velocities. In contrast to these observations, Takei and Holtzman (2009) presented a new theory where they claim a larger effect of a small fraction (~0.1%) of melt for diffusion creep (a factor of 5–10 enhancement of strain rate, still modest compared with the hydrogen effect). However, the validity of their theory is not clear because there is no clear experimental support. (Using an analog material, McCarthy et al. (2011) reported that a small amount (~0.1% or less) of melt has a large effect on deformation. However, the influence of chemical composition is not well corrected in their work and the validity of their conclusion is unclear.) And the stress state calculated in their theory has singularities at grain boundaries that are the transient feature and does not correspond to the true steady-state feature according to the theory by Raj and Ashby (1971). In fact, Raj (1975) and Lifshitz and Shikin (1965) explained transient fast strain rate (by a factor of ~5–10) by such a transient stress state. Furthermore, the dominant mechanism of deformation in most of the asthenosphere is dislocation creep as inferred from strong seismic anisotropy (e.g., Karato et al., 2008); in which case, the influence of partial melt will be even smaller.

In summary, I conclude that although partial melting is ubiquitous in the Earth's mantle due to the presence of volatile elements, partial melting is unlikely to affect physical properties (seismic wave velocities, electrical conductivity, and rheological properties) so much because melt does not completely wet grain boundaries and the melt fraction is small (<0.1%). A region above the 410 km discontinuity is an exception where a substantial reduction in seismic wave velocities is expected (see Coutier and Revenaugh, 2006, 2007; Gao et al., 2006; Tauzin et al., 2010). It is possible that this region has a low viscosity that might decouple convection current across the 410 km discontinuity.

9.05.4 Distribution of Volatile Elements on Earth

A large amount of volatile elements on Earth occurs on its surface (atmosphere, ocean, and sedimentary rocks), but there must be some volatile elements in its deep interior (mantle and core) as seen by the eruption of volcanic gases. The total amount of volatile elements on Earth and their spatial distribution are the key constraints in developing a model of volatile circulation and the evolution of the ocean (Section 9.05.6.3). In this section, I will provide a review on the inferences of volatile distribution.

9.05.4.1 Geochemical Inference of Water and Other Volatile Element Distribution on Earth

The distribution of volatile elements such as hydrogen and carbon can be inferred from the samples from the Earth's interior. These samples include volcanic gases and rock samples (e.g., Bell and Rossman, 1992; Dixon et al., 2002; Ito et al., 1983; Marty, 2012; Peslier, 2010).

Two approaches are used to infer the distribution of volatiles on Earth from rock samples (Figure 19). One is to measure the volatile content in mantle or crustal rock specimens (e.g., Bell and Rossman, 1992; Ingrin and Skogby, 2000; Peslier, 2010). This is a direct approach. In many cases, the temperature and pressure conditions at which a given sample established the last chemical equilibrium can be inferred. Consequently, one can obtain some ideas about the depth variation in the volatile content from these samples (e.g., Peslier et al., 2010). However, there are two limitations with this approach. First, the range of planetary interior that one can explore by this approach is limited by the availability of samples. In the Earth, most of mantle rocks are from depths shallower than ~200 km. Occasionally, some deep samples coming from the transition zone or deeper are identified (Haggerty and Sautter, 1990; Harte, 2010; Walter et al., 2011), but these are very rare (recently, Pearson et al. (2014) reported a high water content (~1 wt%) of ringwoodite (a transition zone mineral) in one of the diamond inclusions, but the relevance of such an observation to the water distribution at the global scale is unclear). Second, even if one can determine the water content from these limited samples, the significance of such results on the water content in the mantle is unclear. This is because hydrogen content in these samples is likely modified (either hydrogen loss or hydrogen gain) during their transport to the surface. In much of the published literature, only hydrogen loss is examined (e.g., Demouchy et al., 2006), but hydrogen (or other volatile elements) may also be gained from the volatile-rich kimberlite magma during the transport of these samples (e.g., Karato, 2010b).

Alternatively, hydrogen (and other volatile) distribution in the mantle can be estimated from the concentration of these elements in the basalts. This is an indirect method for inferring the water content in the mantle because basalts themselves are not mantle rocks. However, this is an important approach

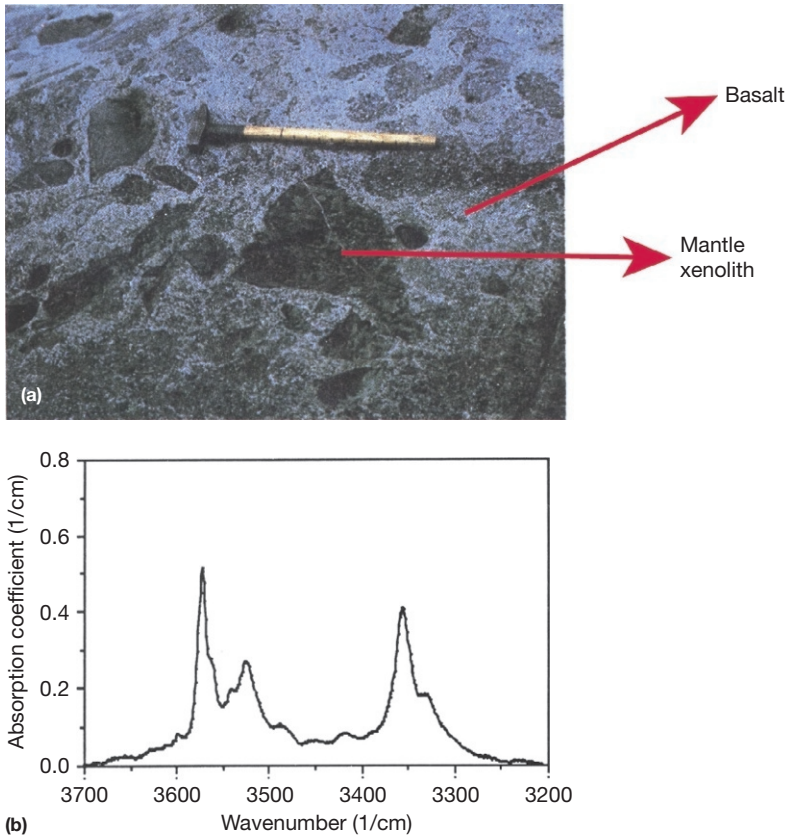


Figure 19 Geochemical (geologic) approach for estimating water (volatile) distribution. (a) A rock specimen containing both basalt and mantle xenoliths. (b) An example of infrared absorption spectrum from which water content in a sample can be estimated. The distribution of volatile elements is to measure the concentration of volatile elements from rock samples (geochemical approach). This method provides direct constraints on the volatile content, but the regions that one can explore is limited and often poorly constrained. Also, the volatile content of rock specimens collected from the Earth's surface may not necessarily represent the volatile content in the Earth's interior. One needs a careful examination of the influence of transport processes of rocks on the volatile content to infer the volatile content in the Earth's interior.

because basaltic magmas are generated at the global scale by partial melting of mantle materials, and hence, their composition reflects mantle composition at a broad space scale (e.g., Hofmann, 1997). In particular, mid-ocean ridge basalt (MORB) is formed by the partial melting of the asthenosphere materials and its chemical composition is nearly homogeneous (Hofmann, 1997). Also, the degree of melting associated with MORB formation (~10%) and hydrogen partitioning between basaltic melt and upper mantle minerals are well known (e.g., Hauri et al., 2006). Consequently, the hydrogen content in MORB provides a robust estimate of water content in the asthenosphere. In contrast, the interpretation of water content of ocean island basalt (OIB) is more complicated because the source regions and the processes of partial melting of OIB are not well understood (e.g., Hofmann, 1997).

Ito et al. (1983) and Dixon et al. (2002) determined the water content in various basalts and inferred the water content in their source regions (**Table 2(a)**). The source regions of MORB, that is, the asthenosphere of the Earth, contain ~0.01 wt% of water (with a factor of ~2 uncertainties). This corresponds to ~4% of the ocean mass, and if this concentration were homogeneous throughout the mantle, the total water content would be ~30% of the ocean mass. The source regions of OIB have a variety of water content, but a common component of OIB source regions, referred to as FOZO (Focus Zone), contains much higher water content (by a factor of ~5–10) than the source region of MORB. This suggests that the deep mantle contains higher water content than the upper mantle. However, because the source regions (particularly their volume) of OIB are unknown, both the total amount of water in the mantle and its distribution are unconstrained from such a geochemical approach. For example, Rüpke et al. (2006) considered that the volume of water-rich FOZO is only ~7% of the mantle, leading to a relatively small total water content of the mantle (~30% of ocean mass). In their model, the mantle is treated as a single component, and therefore, the volume of OIB that has different water content than the MORB source region cannot be large (otherwise, its assumption would be invalidated). Korenaga (2008) obtained a similar conclusion where he assumed that the whole mantle has the same composition as the asthenosphere. A small total water content (also a small content of radioactive elements, i.e., a small Urey ratio (the ratio of heat generation to heat loss of the Earth)) in the mantle in these models is a consequence of the assumption of homogeneous (not layered) mantle chemistry and not a conclusion required from the observations. However, the validity

Table 2 Geochemical estimates of volatile contents[a]

(a) Water contents in the various regions of the Earth's mantle inferred from basalts (after Dixon et al., 2002)

	Depleted mantle[b] (ppm wt)	*FOZO[b] (ppm wt)*	*EM[b] (ppm wt)*
H_2O	100	750	<400

(b) A comparison between the volatile content of the Earth's interior and other objects in the solar system (after Marty, 2012)

	Solar (mol g^{-1})	*Carbonaceous chondrite (mol g^{-1})*	*Depleted mantle (mol g^{-1})*	*Depleted mantle (ppm wt)*	*Bulk Earth (mol g^{-1})*	*Bulk Earth (ppm wt)*
^{12}C	8.6×10^{-11}	2.9×10^{-3}	1.7×10^{-6}	14	6.4×10^{-5}	530
^{14}N	1.8×10^{-11}	1.1×10^{-4}	6.4×10^{-9}	0.12	9.0×10^{-8}	1.7
H_2O	1.2×10^{-10}	6.6×10^{-3}	8.3×10^{-6}	110	2.0×10^{-4}	2700
^{36}Ar	6.3×10^{-13}	3.5×10^{-11}	2.7×10^{-15}	10^{-7}	7.8×10^{-14}	3×10^{-6}

[a]Uncertainties are typically ~50%.
[b]'Depleted mantle' is the upper mantle (source region of mid-ocean ridge basalt (MORB)) that is depleted with volatiles. FOZO is a common component of the source regions of ocean island basalt (OIB). These regions are likely located in the deep mantle. EM is an 'enriched mantle' that is a source region of some of the OIB with large concentrations of various incompatible elements.

of such an assumption is not clear, and there are some observations that suggest that this assumption is not correct, as I will discuss later in the text.

Marty (2012) used a somewhat different approach in which he used the ratio of the concentrations of volatile elements to that of ^{40}Ar in the basaltic rocks (and other samples) (**Table 2(b)**). ^{40}Ar is a highly volatile element produced by the radioactive decay of ^{40}K, and there is essentially no ^{40}Ar in the primordial atmosphere and mantle. Also, as shown in **Figure 8** (also **Figure 9**), ^{40}Ar has relatively high diffusion coefficient, and therefore, most of ^{40}Ar will be removed from solid by partial melting together with volatile elements that in general have high diffusion coefficients (this is not the case for other incompatible elements such as Ce that have low diffusion coefficient). (The H/Ce ratio is often assumed to be a constant and used to infer the hydrogen content (e.g., Dixon et al., 2002). If Ce is not completely removed from minerals by partial melting due to slow diffusion (**Figure 8**), such a method introduces a systematic bias. Also, there is no guarantee that the partitioning coefficients of H and Ce are the same for various pressure and temperature conditions.)) Consequently, when a rock is degassed near the surface, almost all ^{40}Ar must be degassed from the original rock similar to all other volatile elements (H, C, N, Ne, etc.). Therefore, their ratios, for example, $H/^{40}Ar$, reflect their ratios in the mantle. If the concentration of K in the mantle is known, ^{40}Ar content in the mantle can be calculated from the measured ^{40}Ar content in the atmosphere. Marty (2012) used the geoneutrino observations (Gando et al., 2011) to estimate K content and inferred that the Earth's water content is 0.27 ± 0.13 wt% of the total mass, $\sim10\pm5$ ocean mass. The geoneutrino inference of K content has a large uncertainty, but high concentration of K and other volatile elements in the deep mantle was also inferred by Allègre et al. (1996) and Halliday (2013) from geochemical considerations. The higher concentration of incompatible elements such as H and K in the deep mantle than that in the upper mantle is a natural consequence of transition zone partial melting (Bercovici and Karato, 2003).

This value is in good agreement (within a factor of ~2) with the estimate by Karato (2011) using geophysical observations as will be discussed in the next section that shows that the water content in the transition zone and the lower mantle is much larger than that in the upper mantle. However, these estimates differ considerably from those by Hirschmann (2006), Hirschmann and Dasgupta (2009), and Rüpke et al. (2006) who assumed that the water content in the deep mantle is similar to that in the upper mantle except for very small water-rich regions as a source for OIB.

Distribution of other volatile elements such as C (carbon) can also be estimated from rock samples. For example, the carbon content in the mantle was estimated by Dasgupta and Hirschmann (2010), Marty (2012), and Wood et al. (1996) from geochemical observations. In particular, based on the composition of MORB and OIB, Hirschmann and Dasgupta (2009) inferred that H/C ratio in the upper mantle is larger than that in the deep mantle (transition zone and below), implying that the lower mantle has a relatively high carbon content. The abundance of other volatiles such as S, N, Ne, and Ar was also estimated (e.g., Marty, 2012). Marty (2012) showed that the concentrations of most of volatile elements are higher in the deep mantle than those in the upper mantle (depleted mantle), and the total amount of volatile elements in the Earth is on the order of 1–3% of those in carbonaceous chondrite (**Table 2(b)**) (a similar conclusion was obtained by Allègre et al., 1986/1987).

9.05.4.2 Geophysical Inference of Water Distribution

Although the geochemical approach for estimating the volatile distribution is more direct, its major limitations are the difficulties in locating where those volatile elements are present and in estimating the volume of each region ('reservoir') except for the source region of MORB, that is, the asthenosphere. For example, direct sampling from the deep interior of planets is difficult, and even for the Earth, there is little consensus as to the location and the size of the source regions of OIB. In fact,

some assume that the composition of most of the deep mantle is the same as that of the asthenosphere (e.g., Hirschmann, 2006; Korenaga, 2008; Rüpke et al., 2006), whereas there is some geochemical evidence that the deep mantle contains much higher volatile contents (e.g., Allègre et al., 1996; Marty, 2012). As I will discuss in **Section 9.05.5**, these models lead to very different views of global volatile circulation.

In view of these limitations, a geophysical approach, that is, remote sensing, is an alternative approach to be explored (Karato, 2006b; **Figure 20**). Various remote sensing techniques have been developed to explore planetary interiors including (i) geodesy (gravity, shape of planets, and orbit), (ii) seismology (elastic and anelastic properties), and (iii) geomagnetic induction (electrical conductivity). In these methods, one uses the observations obtained outside of a planet that are related to some physical (and chemical) properties of planetary interiors. If one knows a relation between these observable properties and volatile content, then one can infer volatile distribution in a planet. A major advantage of this approach is that in most cases, these measurements provide information of distribution of physical properties in the entire planet and sometimes provide three-dimensional models.

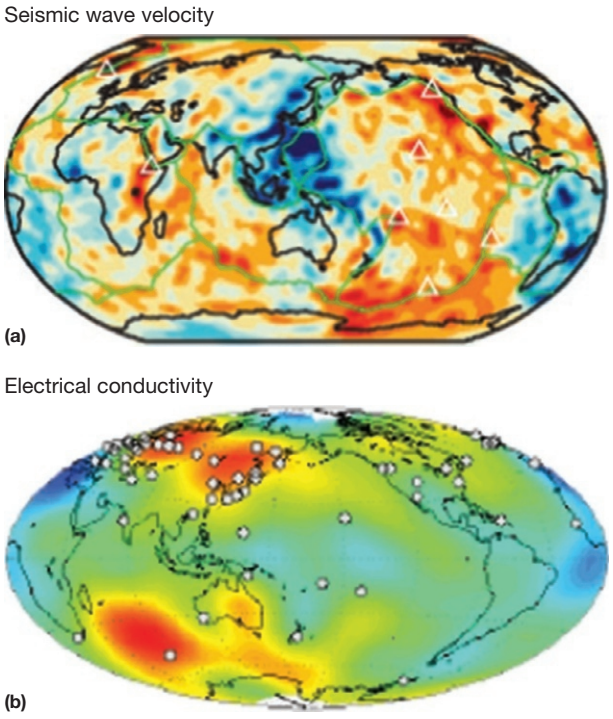


Figure 20 Data that could be used in geophysical remote sensing of volatiles. (a) Anomalies in seismic shear wave velocity in the transition zone (Ritsema et al., 2011). (b) Anomalies in electrical conductivity in the transition zone (Kelbert et al., 2009). The distribution of volatile elements can be inferred from geophysical observations if some geophysically measurable properties are sensitive to volatile content and if geophysical observations with sufficient resolution are available. In order to apply these methods, one needs careful studies on the influence of volatile elements on geophysical properties. Only hydrogen is known to produce its presence in geophysical properties such as electrical conductivity, seismic attenuation, and anisotropy. From this approach, one can investigate the spatial distribution of volatile content better than the geochemical approach.

However, unlike geochemical approach where volatile content is directly measured on rock samples, geophysical approach is indirect. In order to provide useful constraints on the volatile distribution from geophysical approach, one needs to identify which properties one may use and to establish a relationship between the volatile content and these physical properties. Various physical properties depend not only on volatile content but also on other parameters. As I discussed in **Section 9.05.3**, hydrogen is exceptional among various volatile elements in influencing some of the geophysically observable properties to the extent that one can place important constraints on its distribution. This is due to the weak chemical bonding of hydrogen with other atoms such as oxygen and relatively large solubility of hydrogen in most minerals. Therefore, hereafter in this section, I will focus on the estimate of hydrogen content. If the ratio of hydrogen content to the content of other volatile elements is known, then one can use hydrogen as a geophysical tracer to infer the distribution of other elements.

The dependence of a physical property, X, on various parameters may be symbolically written as

$$X = X(C_W, T; \xi) \quad [13]$$

where C_W is the water content, T is the temperature, and ξ are 'other factors' such as the major element chemistry (pressure effect is not included here. If needed, temperature could be interpreted as a homologous temperature, $T/T_m(P)$ ($T_m(P)$, melting temperature), to include the pressure effect). Since all of them can vary in a planetary interior, if a property can be used to infer water content, its sensitivity to 'other factors' (major element chemistry) and temperature must be weak or well characterized. As I will discuss later in the text, some physical properties such as electrical conductivity and seismic attenuation (tidal dissipation) are relatively insensitive to the major element chemistry (or one can infer the influence of major element chemistry with sufficient accuracy). In these cases, one can infer combinations of C_W (water content) and T (temperature) that explain geophysical observations. In a case where temperature can be estimated using other approaches (this is the case for the Earth's mantle), water content can be inferred from such a geophysical observation.

9.05.4.2.1 From seismological observations

Seismology provides high-resolution images of the Earth's interior (there are some data on the Moon as well; Garcia et al., 2011). Among others, seismic wave velocities can be determined with the highest precision and provide the best constraints on the composition of the Earth's interior (e.g., Dziewonski, 2000). Also, in some cases, the depths of seismic discontinuities can be determined that are also sensitive to the water content (e.g., Deuss, 2009; Shearer, 2000). Consequently, there have been some attempts to use seismic wave velocities (e.g., Jacobsen et al., 2004, 2008; Jacobsen and Smyth, 2006) and/or the depth of seismic discontinuities to infer the water content. Particularly interesting is an approach used by Meier et al. (2009) and Suetsugu et al. (2006) where the anomalies in seismic wave velocities and in the depth of

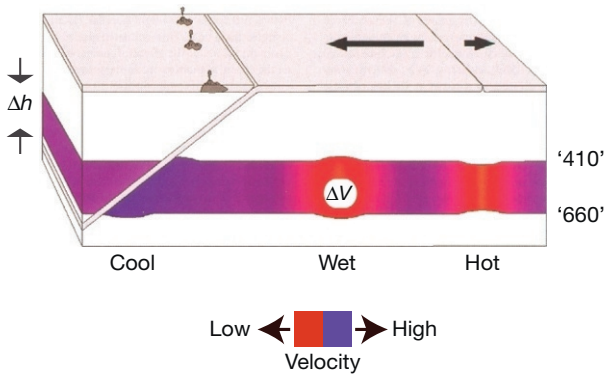


Figure 21 A schematic diagram showing how temperature and water may affect the seismic wave velocity and the topography of the 410 km and 660 km discontinuities (thickness of the transition zone). Temperature and water content affect seismic wave velocity and the topography of seismic discontinuities (or the transition zone thickness) differently. Therefore, by using the observations on the anomalies in seismic wave velocity and transition zone thickness, one can determine the anomalies in temperature and water content separately. However, in order for the method to work, temperature and water content must be the most important variables to affect seismic wave velocities and the topography of discontinuities. If some other variables such as the major element chemistry affect these properties more than the temperature and water content, this method will not work.

seismic discontinuities (thickness of the transition zone) are inverted together for the anomalies in temperature and water content (**Figure 21**):

$$\begin{pmatrix} \dfrac{\Delta V_{P,S}}{V_{P,S}} \\ \dfrac{\Delta h}{h} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \log V_{P,S}}{\partial T} & \dfrac{\partial \log V_{P,S}}{\partial C_W} \\ \dfrac{\partial \log h}{\partial T} & \dfrac{\partial \log h}{\partial C_W} \end{pmatrix} \begin{pmatrix} \Delta T \\ \Delta C_W \end{pmatrix} \qquad [14]$$

where $V_{P,S}$ is the seismic wave velocity, h is the thickness of the transition zone, T is the temperature, C_W is the water content, and ΔX is the anomaly of X. In such a formulation, one has two unknowns (water content and temperature), and using two independent observations (anomalies in seismic wave velocity and anomalies in the depth of seismic discontinuity), one can uniquely determine the temperature and water content.

Although this method would work in principle, such an approach led to puzzling conclusions such as water-poor transition zone in eastern Asia and water-rich transition zone in the central Pacific when it is applied to a global scale (Meier et al., 2009). This suggests that the assumption that the anomalies of seismic wave velocity and the topography of discontinuity (i.e., the thickness of the transition zone) depend primarily on water content and temperature is not appropriate. As seen from **Figures 14** and **15**, both seismic wave velocity and the topography of the seismic discontinuity depend only weakly on water content under typical (hot) mantle conditions. In other words, in the case of seismic wave velocity and the thickness of the transition zone, the influence of ξ in eqn [14] is so strong that the variation in $X(C_W, T; \xi)$ is mostly caused by that of ξ (major element chemistry) rather than the water content (C_W). In these cases, rather than the anomalies in the water content, ΔC_W, one should use another variable such as the variation in major element chemistry and temperature (ξ, T) as unknowns. Even in these cases, water content could be inferred if major element chemistry is very precisely known. This is almost impossible because in order to detect ~0.1 wt% of water, one needs to know the iron content better than ~0.1%, for example. I conclude that it is difficult to infer water content from seismic wave velocity and/or the topography of the mantle discontinuities.

There are other types of seismological observations that provide better constraints on the water content. The first is seismic wave attenuation and the second is seismic anisotropy. When seismic wave attenuation is well resolved, the observed attenuation can be used to infer water content (and temperature). Such an attempt was made for the Earth by Shito et al. (2006) who found water-rich regions in the deep (~300 km) wedge mantle in the western Pacific. Similarly, Karato (2013a) applied such a method to the Moon using the inferred tidal dissipation and electrical conductivity as constraints. He found evidence of a substantial amount of water in the deep lunar mantle similar to the upper mantle of the Earth. The limitations with this approach are that the relation between seismic wave attenuation (or tidal dissipation) and water content is not precisely constrained and that geophysical observations on attenuation have large uncertainties (e.g., Dalton et al., 2009; Romanowicz and Durek, 2000). In contrast, tidal dissipation is determined by the geodetic measurements of the shape of planets and can be determined precisely (errors less than ~10%) (e.g., Ray et al., 2001; Williams et al., 2001). However, tidal dissipation is a property of the whole planet, and one needs a model for the depth-dependent attenuation to interpret tidal dissipation (Karato, 2013a).

Deformation microstructures such as the LPO of olivine depend on the water content and other parameters such as temperature and stress (**Section 9.05.3.3**; **Figure 16**; Karato et al., 2008). Consequently, for a given flow geometry, seismic anisotropy depends on the water content (and other factors). Therefore, when flow geometry and other factors such as temperature are well known, one can infer the distribution of water content from seismic anisotropy. For example, anomalous seismic anisotropy in the central Pacific (Ekström and Dziewonski, 1998) was explained by the redistribution of water caused by deep partial melting associated with the Hawaiian plume (Karato, 2008b). Kneller et al. (2005) explained the complex pattern of anisotropy in the wedge mantle (mantle above a subducting slab) by the variation in LPO caused by the variation in stress, temperature, and water content. An obvious limitation of this approach is that the location of seismic anisotropy is not easily constrained particularly when one uses the body waves (e.g., SKS; Savage, 1999; Silver, 1996). Also, some laboratory studies suggest more complex behavior of LPO in olivine (e.g., Ohuchi and Irifune, 2013).

9.05.4.2.2 From electrical conductivity

Electrical conductivity in a planet can be inferred from the observations of electromagnetic induction (e.g., Banks, 1969; Hood et al., 1982; Kelbert et al., 2009; Rikitake, 1966). Temporal variations of an external magnetic field induce electrical currents in a planet that in turn induce magnetic field. By measuring the electromagnetic field on the planetary surface, one can determine the ratio of the external and internal field. The induced current is proportional to electrical conductivity,

and hence by determining the ratio of induced field to external field, one can infer the electrical conductivity. The depth variation of electrical conductivity can be inferred from the frequency dependence of electromagnetic induction. In many cases, the lateral variation in electrical conductivity is large comparable with the depth variation (e.g., Baba et al., 2010; Banks, 1969; Evans et al., 2011; Hood et al., 1982). This suggests the strong sensitivity of electrical conductivity to some parameters that change laterally such as temperature and the water content.

The resolution of electrical conductivity estimate is not as high as that of seismological observations. However, it is important to understand that the range of the water content we may have in a terrestrial planet is broad, from ~1 ppm wt or less (nearly completely dry) to ~1 wt% (nearly saturated). Consequently, if one can infer water content within one order of magnitude, this will be an important progress. Electrical conductivity observations have enough resolution for this purpose.

Figure 22 compares the conductivity–depth profile models corresponding to various water contents with geophysically inferred conductivity–depth profiles (Karato, 2011). Such a comparison suggests strongly layered water content in the Earth's mantle, ~0.01 wt% in the upper mantle (asthenosphere) and ~0.1 wt% in the transition zone (Figure 23). These results also show a large regional variation in the water content: a high water content in the transition zone in East Asia (Ichiki et al., 2006), a low water content in the transition zone in eastern/southern Europe (Tarits et al., 2004), and a low water content in the asthenosphere in the western Pacific (Baba et al., 2010). Currently, there are no constraints on the water content of the lower mantle because there are no laboratory data on the influence of the water content on the electrical conductivity of lower mantle minerals.

A similar study was made for the Moon where geophysically inferred electrical conductivity–depth profile was interpreted by a combination of temperature–water content (Karato, 2013a). Combining the constraints from tidal dissipation, which also provides a constraint on the temperature–water content combination, with electrical conductivity, Karato (2013a) estimated the temperature and water content separately and concluded that the presence of water similar to the upper mantle of the Earth (~0.01–0.001 wt%) is needed to explain these observations.

9.05.4.2.3 From rheological properties

Given the sensitivity of rheological properties to water content summarized in Section 9.05.3.2, one might consider that the variation of rheological properties (viscosity) of the mantle might be used to infer the water content. Such an approach is, however, hampered by the poor spatial resolution of the geophysical estimate of rheological properties and by the strong trade-off between the water and temperature effects. For example, Cadek and Fleitout (2003) and Nakada and Lambeck (1991) presented evidence for the lateral variation in viscosity, and Dixon et al. (2004) interpreted some of geophysical observations in terms of the lateral variation in water content. However, the results of these studies are subject to large uncertainties due to the strong trade-off between the temperature and water effects.

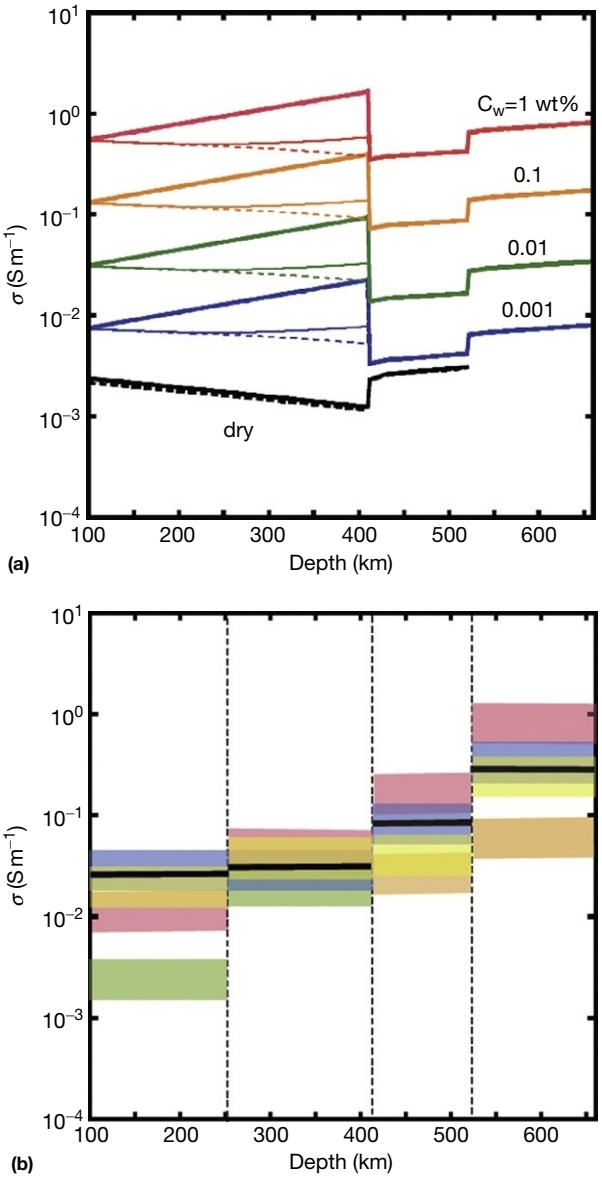


Figure 22 (a) Models of electrical conductivity versus depth relations in the Earth's upper mantle and transition zone (Karato, 2011). Water content is assumed to be independent of depth. The depth variation in conductivity is caused mainly by depth variation in temperature and mineralogy (phase transformations). (b) Geophysically inferred electrical conductivity in the upper mantle and transition zone (Karato, 2011). Different colors correspond to the conductivity in different regions (for details, see Karato, 2011).

9.05.4.3 Other Volatile Elements

Similar to hydrogen, the concentration of carbon in the mantle can be inferred from the composition of basaltic rocks (Dasgupta and Hirschmann, 2010; Hirschmann and Dasgupta, 2009; Wood et al., 1996). Both hydrogen and carbon are modest to highly incompatible elements, and a majority of them move to melts upon partial melting. However, there are two important differences in the behavior of carbon and hydrogen upon partial melting. First, carbon is more incompatible than

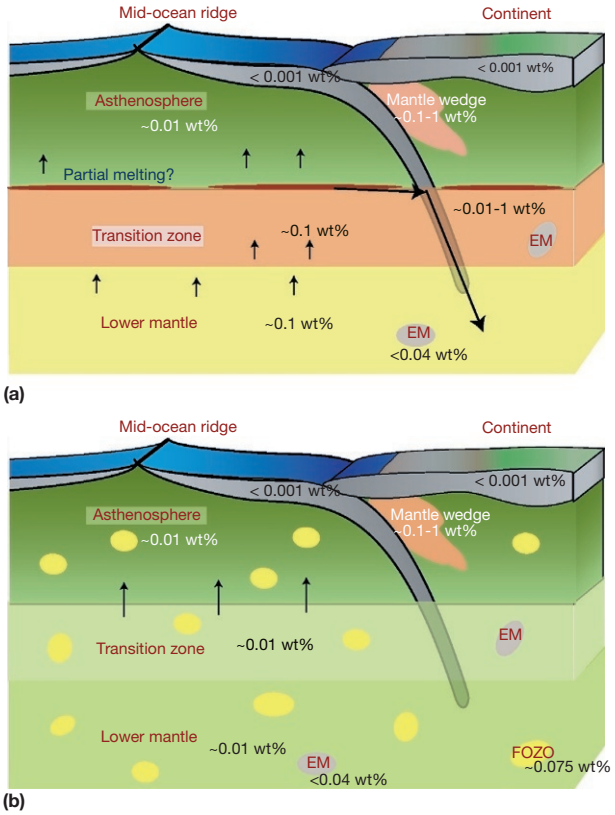


Figure 23 Models of water distribution in the Earth's mantle. (a) A layered water content model. Most geophysical models to explain electromagnetic induction observations require a substantial increase in electrical conductivity at ~410 km. This implies the increase in water content across the 410 km discontinuity. This model is consistent with the geochemical observations on the rare gases (Marty, 2012). (b) A raisin bread model. A model of patchy water-rich regions in the deep mantle proposed by Rüpke et al. (2006) and Korenaga (2008) is not consistent with the electrical conductivity and the rare gas observations (Marty, 2012). Reproduced from Karato S (2011) Water distribution across the mantle transition zone and its implications for global material circulation. *Earth and Planetary Science Letters* 301: 413–423.

hydrogen (Shcheka et al., 2006). Second, the diffusion coefficient of carbon in minerals is generally much slower than that of hydrogen. These two factors make carbon depletion more efficient upon partial melting. If indeed the upper mantle (asthenosphere) were formed as a result of partial melting at ~410 km, then the C/H ratio would be higher in the transition zone (and lower mantle) than that in the upper mantle.

Using the carbon and hydrogen contents in MORB and OIB, Hirschmann and Dasgupta (2009) inferred that the H/C ratio of the deep mantle (transition zone and the lower mantle) is substantially smaller than that in the upper mantle. The inferred difference in the H/C ratio between the upper mantle and the deep mantle is difficult to explain if there is no major mass segregation in between these two regions. However, because of the different behaviors of C and H upon partial melting, the inferred variation in the H/C ratio is naturally attributed to the effect of partial melting and melt-solid segregation at (near) 410 km. The abundance of other volatile elements is reviewed by Marty (2012) (**Table 2(b)**).

9.05.4.4 Water Content in Other Planetary Bodies

The water contents in some meteorites were measured and the water contents of comets were estimated (Jarosewich, 1990; Marty and Yokochi, 2006). Comets are volatile-rich (~50 wt%), so are carbonaceous chondrites (~10 wt%). Ordinary chondrites have ~0.1 wt% of water. The most depleted meteorite is enstatite chondrite that has less than 0.1 wt% of water. Most of ordinary chondritic meteorites are considered to come from the inner asteroid belt (2.1–2.5 AU; AU, the astronomical unit (the distance between the Earth and the Sun), 1.50×10^8 km from the Sun), whereas carbonaceous chondrites are from the outer region of the asteroid belt (~3 AU) (Gradie and Tedesco, 1982). This suggests that materials within the snow line (~2.7 AU; see **Section 9.05.5.1.2**) have enough water to be an important, perhaps dominant, source of the Earth's water.

Water contents of the Moon were inferred from the lunar samples, resulting in a wide range of values. Among them, the most robust is the measurements on basaltic inclusions in olivine that showed the water content similar to that in mid-ocean ridge basalt (Hauri et al., 2011) (see also Hui et al., 2013; Saal et al., 2013). Geophysical observations suggesting the Earth-like water content in the lunar mantle are reviewed by Karato (2013a). He used the geophysically inferred electrical conductivity–depth profile and the tidal dissipation to infer the water content (and temperature) to conclude that the water content of the lunar mantle is similar to that of the upper mantle of the Earth, a conclusion similar to that obtained

from the geochemical study by Hauri et al. (2011). Saal et al. (2013) also provided isotopic observations suggesting that water in the Moon is indigenous and the water and other volatiles in the Moon likely have similar origin as those in the Earth.

Water content in the Martian mantle can be inferred from the composition of the Martian meteorites (e.g., McSween et al., 2001; Mysen et al., 1998). However, the interpretation of observed water contents of these samples is not unique and different models were proposed (McSween et al., 2001; Mysen et al., 1998). Similar to the Moon, geophysical observations would provide additional constraints on the water content in the Martian mantle. The tidal dissipation of Mars was determined from the analysis of the orbits of its satellites (Lainey et al., 2007). If the electrical conductivity of Mars were inferred from electromagnetic induction measurements, it would be possible to infer the water distribution in Mars. For Venus, no data on tidal dissipation nor electrical conductivity are available. It is often discussed that the Venusian mantle is likely water-poor based on the observed positive correlation between topography and gravity (Kaula, 1990) and the composition of the atmosphere (e.g., Chapter 4 of de Pater and Lissauer, 2010). However, there is some suggestion that Venus had a large amount of hydrogen in the past (e.g., Donahue et al., 1982; Grinspoon, 1987; Kasting and Pollack, 1983). Because plate tectonics does not operate on Venus at least in the recent ~500 Myr, degassing is less efficient and Venusian mantle might contain a substantial amount of water (see also Halliday, 2013).

9.05.5 Volatiles During Planetary Formation

Planets are formed as a by-product of star formation, which is caused by the gravitational collapse of a molecular cloud (Shu et al., 1987). By gravitational collapse, such a cloud will be heated up (these materials will also be heated up by the radiation of the mother star). Most of the materials of the molecular cloud go to the center of the cloud from which a star is formed. However, there will be some leftover materials, from which planets are formed. The composition of these materials is initially similar to that of the molecular cloud from which the star was formed. In fact, recent search for exoplanets shows that a large fraction of the stars (>40%) have planets (Howard et al., 2010).

During the cooling of the nebula associated with the strong radiation generated by the early-stage star ('T Tauri' star, a star in the later stage of the Hayashi contraction phase; Hayashi, 1961), most of the volatile components in the gas will be blown off to define the initial composition of a terrestrial planet. In the later stage of planetary formation, extensive heating will occur by collisions that can result in the volatile loss. (In a giant impact, most materials are gravitationally bound and not much volatile loss will occur. However, if temperature exceeds the vaporization temperature, then planets (satellites) formed from the once vaporized materials will be depleted of volatiles to some extent during condensation (see, e.g., Karato, 2013b).) Collisions also lead to orbital scattering that enhances the mixing of materials with different volatile contents (e.g., Kokubo and Ida, 2000; Morbidelli et al., 2000). And finally, in the late stage of planetary formation, magma ocean and atmosphere will be formed that redistribute volatile elements (e.g., Abe and Matsui, 1985; Abe et al., 2000; Elkins-Tanton and Grove, 2011). After a planet is formed, some volatiles will be lost through the outer layer of the atmosphere (e.g., Abe et al., 2011; Hunten et al., 1989), but there will be a long-term circulation of volatiles involving degassing and regassing associated with mantle convection and core cooling (e.g., Franck and Bounama, 2001; McGovern and Schubert, 1989; Rüpke et al., 2006). The timescales of these processes are summarized in **Table 3**, showing that most of these stages have markedly different timescales so that some of them can be discussed separately. However, there are a few cases where the mutual feedback of two processes may be important including the formation of a satellite associated with a giant impact.

Table 3 Timescale of various planetary processes

Process	*Timescale (Myr)*
Nebular collapse, cooling	$\sim 10^{-4}$–10^{-3}
T Tauri stage (Hayashi contraction phase)	$\sim 10^{-2}$–1
Planetary formation	~10–10^{2}
Planetary evolution	~10^{3}–10^{4}

9.05.5.1 Acquisition and/or Loss of Volatile Elements in the Early Stage of Planetary Formation

9.05.5.1.1 General introduction

Terrestrial planets are made of condensed materials. Therefore, condensation from a hot gas is the first stage by which the volatile content of terrestrial planet is controlled. Upon cooling, refractory materials are condensed in the early hot stage and volatile materials are condensed later in the cool stage. The condensed materials not only will sink to the equatorial plane of the nebula and eventually form planets but also may migrate toward or away from the star. When gaseous phases are blown off by the strong solar wind at the T Tauri stage of the mother star, condensation will cease and the condensed materials at this stage will become the source materials of terrestrial planets (strong radiation will blow off most of the solar nebula composition gas from terrestrial planets, but not from giant planets because of the strong gravity for the latter). (Large impact could also blow off the preexisting atmosphere, but the timing of a large impact is 10 s of Myr after the formation of the Sun, and therefore, it will be after the blowoff of the primary atmosphere by the solar radiation during the T Tauri stage.)

Assuming this scenario, the composition of the condensed materials that will become terrestrial planets can be calculated from the temperature of the nebula at the time of the T Tauri phase. Whenever the timing of the T Tauri stage is, the temperature in the nebula at any distance from the mother star decreases with distance. Consequently, there will be a gradient in chemical composition of the condensed matter. Lewis (1972) proposed that the composition of the terrestrial planets can be explained by the condensation temperatures of various materials. Similarly, Gradie and Tedesco (1982) reported that the composition of asteroids inferred from the reflection

spectroscopy varies systematically as a function of the distance from the Sun. This is consistent with the condensation model. They also discussed that this observation suggests that there is not much orbital scattering of these bodies after their formation. However, recently, DeMeo and Carry (2014) showed a more complex compositional distribution in the asteroid belt, suggesting the importance of mixing. Also, Cassen (1996) developed a model to explain the variation of moderately volatile elements in some carbonaceous meteorites using such a model where radial transport of matter is also considered.

9.05.5.1.2 The snow line

There will be a critical distance from the mother star at which temperature becomes cooler than the condensation temperature of ice when the mother star is at the T Tauri stage of evolution. This critical distance (line) is called the snow line (Hayashi, 1981; or the 'frost line' or the 'ice line'). Hayashi (1981) estimated that the location of the snow line for the solar nebula is ~2.7 AU (uncertainties in such an estimate will be discussed later). Beyond the snow line, it is often argued that a large amount of volatile-rich materials is condensed that makes it possible to nucleate a core quickly on which a giant planet could be formed. (The formation of giant planets also needs a large amount of materials. Therefore, the distribution of giant planets is also controlled by the distribution of materials (the surface density) in the solar nebula.) The concept of the snow line provides a natural explanation of the presence of two classes of planets: terrestrial planets within the snow line (<2 AU) and gaseous planets away from the snow line (>4 AU).

It is often assumed that the condensed materials within the snow line are completely dry. If this were true, then in order for terrestrial planets formed within the snow line (e.g., the Earth) to acquire any water (hydrogen), substantial orbital scattering to increase the eccentricity would be needed (e.g., Albaréde, 2009; Morbidelli et al., 2000; Raymond et al., 2007). Because the orbital scattering becomes large only in the later stage of planetary accretion (Kokubo and Ida, 2000; Morbidelli et al., 2000), such a model would imply that water on Earth was acquired in the late stage of its accretion, say, ~50–100 Myr after the formation of the solar nebula (Albaréde, 2009) when the core was already formed (~30 Myr after the formation of the Sun; Kleine and Rudge, 2011). Wood et al. (2010), however, presented geochemical observations (abundance of moderately volatile and siderophile elements) suggesting that a substantial amount of volatile elements was present on Earth at the time of core formation. If most of volatile elements such as H were acquired after core formation, the core would not have much volatile elements. In contrast, if a substantial amount of volatile elements were acquired before core formation, then the core will have some volatile elements.

The location of the snow line depends on the opacity of the nebula and the radiation from the star at the time of the T Tauri stage of a star, both of which have large uncertainties. First, the temperature distribution in the nebula depends strongly on the opacity that in turn depends on poorly constrained amount of condensed matter (Cameron and Pine, 1973). In fact, Sasselov and Lecar (2000) proposed a model where dust grains play an important role in the thermal structure and suggested that the snow line is at ~1 AU for the solar system. Second, the condensation temperature of ice depends on the total pressure of the nebula that is model-dependent (see eqn [2]). The total pressure depends on the mass of the nebula and is different between the minimum mass model of Hayashi et al. (1985) and a model by Cameron (1988) where the mass of the nebula is considered to be the same as the mass of the Sun. This leads to a large uncertainty in the condensation temperature (see eqn [2]) and hence in the locations of the snow line.

9.05.5.1.3 Water (hydrogen) within the snow line?

When the concept of the snow line is used in its simplest form, the formation of hydrous minerals such as serpentine is not considered based on the simple collision theory of the kinetics of chemical reaction (e.g., Fegley, 2000). However, based on some experimental observations, Ganguly and Bose (1995) pointed out that the kinetics of formation of hydrous minerals might be much faster (more than 10^{14} times of magnitude faster!) than Fegley (2000) calculated (see also Jug et al. (2007)and Mejias et al. (1999) for the fast kinetics of hydrous mineral formation). If these experimental observations can be applied to the solar system, then the formation of hydrous minerals within the snow line would be possible. For these reasons, the concept of the snow line should be used with caution in the discussion of volatile acquisition in a growing terrestrial planet.

Another possibility was proposed by Muralidharan et al. (2008) who suggested that the surface adsorption of volatiles on dust grains might contribute to the volatile budget of terrestrial planets. In such a case, even materials within the ice line will acquire some volatiles. However, in order to explain an appreciable amount of volatiles by this mechanism, they need to assume large surface area caused by irregular morphology of grains. The validity of such an assumption is unclear. Recently, Vattuone et al. (2013) reported a more detailed experimental and theoretical study on water adsorption. Again, however, the amount of water that could be kept in these materials by adsorption depends strongly on the size of grains that is not well constrained.

The observed finite amount of water in chondrite suggests that there is some water within the snow line although the processes by which such a finite amount of water (and other volatiles) was acquired are not well understood.

9.05.5.1.4 Gas → solid versus gas → liquid condensation

Condensation occurs during nearly isobaric cooling (small reduction in pressure occurs during cooling). Therefore, whether gas → solid or gas → liquid condensation occurs is essentially determined by the relative position of the triple point pressure P_c and the actual total pressure (**Figure 3**).

Most previous discussions on the condensation of volatile elements in cosmochemistry have been made assuming the gas ↔ solid reactions (vaporization and condensation; Grossman and Larimer, 1974). As shown in **Figure 3**, such reactions are relevant at low pressures. In most of the environments in the primitive solar nebula, the gas pressure is low (Cameron and Pine, 1973; Hayashi, 1981) and such a model is appropriate. However, under some conditions where the total pressure was high, liquids will play an important role.

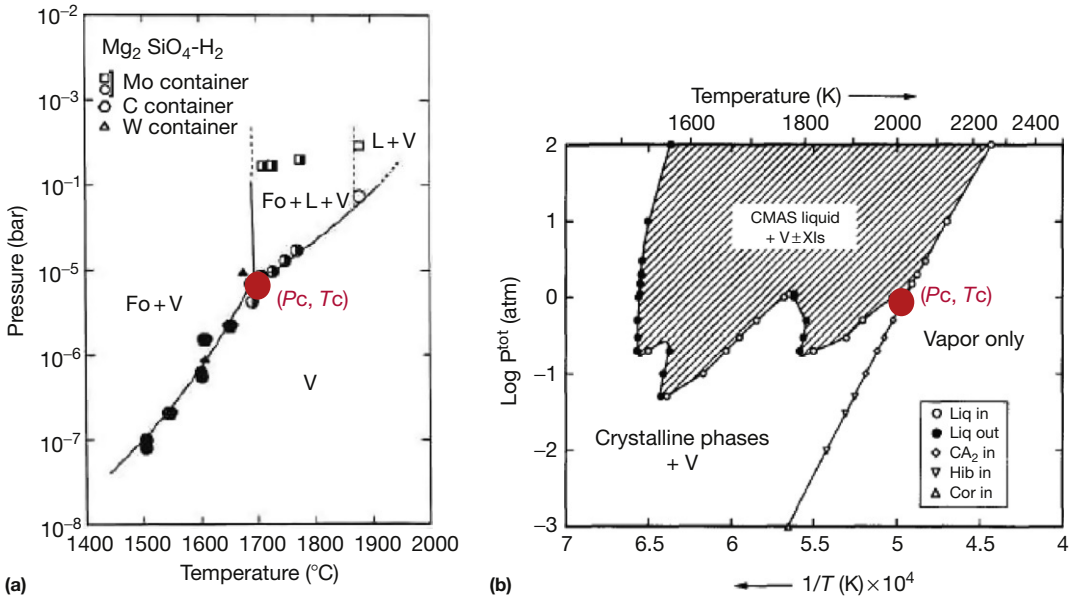


Figure 24 Phase diagrams of two systems in the cosmochemical environment. (a) The H_2–Mg_2SiO_4 system (Mysen and Kushiro, 1988). (b) The CaO–MgO–Al_2O_3–SiO_2+H_2 and other volatile systems (Yoneda and Grossman, 1995) (a substantial fraction of refractory elements is in the dusts). Condensation occurs during cooling that is nearly isobaric (pressure decreases slightly with cooling). For the solar nebula (low pressures), condensation occurs as gas → solid, whereas for a lunar disk (high pressures), it occurs as gas → liquid. A change in composition during condensation may also promote condensation of liquid phases (e.g., Ebel and Grossman, 2000).

Figure 24 shows phase diagrams relevant to planetary formation through condensation. Condensation occurs through (nearly) isobaric cooling. Consequently, the sequence of condensation depends strongly on the relative value of the triple point pressure P_c and the total pressure of the system. If the total pressure exceeds P_c, then condensation forms liquids first. This critical pressure (and temperature) depends on the composition of the system. P_c of a system H_2 and Mg_2SiO_4 (Mysen and Kushiro, 1988) is smaller than that of a system CaO–MgO–Al_2O_3–SiO_2+hydrogen and other volatiles where a substantial fraction of refractory elements are in the dusts (Yoneda and Grossman, 1995). The study of Mysen and Kushiro (1988) is an experimental study, whereas that of Yoneda and Grossman (2005) is a model. (The difference may also be due to the uncertainties in the model calculation by Yoneda and Grossman (2005).) In the solar nebula, a typical pressure is ~10–10^2 Pa, whereas in the Moon-forming disk, the pressure is much higher (10^4–10^7 Pa; Karato, 2013b). Condensation under relatively high pressure may also occur during the collisions of planetesimals as inferred from the chemistry of some meteorites (e.g., Campbell et al., 2002). Also, cooling after heating by a shock wave will involve gas → liquid condensation that is a possible mechanism of chondrule formation (e.g., Galy et al., 2000).

The gas → liquid condensation is important in relation to the volatile acquisition in a growing terrestrial planet. When only the gas → solid condensation is considered, what matters most is the condensation of volatile-rich minerals such as ice. However, when liquid phases are formed by condensation, they can store a much larger amount of volatiles than solids (Figure 25). Karato (2013a) explained the inferred high water content of the Moon (similar to that of the Earth's upper mantle) by the condensation to liquid phases and subsequent rapid accretion.

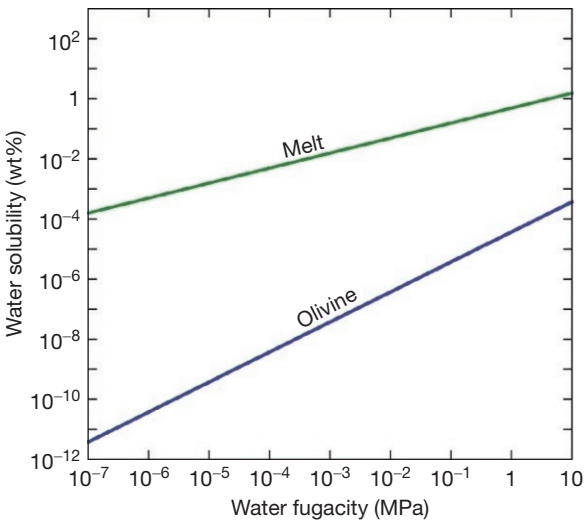


Figure 25 A comparison of solubility of hydrogen in melt (basalt) and solid (olivine). Hydrogen solubility in both basalt (silicate melt) and olivine increases with water fugacity (pressure). However, hydrogen solubility in basalt is much higher than that in olivine in all geologically plausible conditions.

9.05.5.2 Behavior of Volatiles During the Later Stage of Planetary Formation

9.05.5.2.1 Collision-induced melting and vaporization

Terrestrial planets grow as a result of collisions of small planetary bodies ('planetesimals') formed after condensation (e.g., Safronov, 1972). Once some volatile elements are acquired in the condensed matter (planetesimals) during the cooling of

the nebula, the next process that will redistribute volatile elements is the collision of planetesimals. The current models of planetary formation suggest that the average size of planetesimals increases with time, and in the later stage of planetary formation, relatively large bodies collide (e.g., Wetherill, 1990). Collisions of large bodies release a large amount of energy during a short time and hence produce high degree of heating and resultant redistribution of volatile elements.

There have been experimental studies (e.g., Lange and Ahrens, 1982; Tyburczy et al., 1986, 2001) as well as theoretical modeling (e.g., Ahrens and O'Keefe, 1972; Canup, 2004b; Pritchard and Stevenson, 2000; Stevenson, 1987) on these processes. The essential cause of impact-induced melting and/or vaporization is heating by the release of gravitational energy. Therefore, the degree of heating increases with the size of colliding (and collided) materials. The *maximum* (assuming no kinetic energy when bodies are far apart) temperature at the surface of a growing planet can be calculated assuming that all the gravitational energy is converted to heat, that is,

$$(\Delta T)_{\max} = \frac{4\pi R_{\mathrm{P}}^2 G\rho}{3C_{\mathrm{P}}} \qquad [15]$$

where G is the gravitational constant, R_{P} is the radius of a planet, ρ is the (average) density of a planet, and C_{P} is the specific heat. For the melting temperature of typical silicates at low pressure (~1500 K), one finds that the radius of $R_{\mathrm{P}} > 2000$ km will be needed. Therefore, in the later stage of planetary accretion, the surface of a growing planet will be melted if the size exceeds ~1/3 of the Earth and if the gravitational energy is efficiently converted to heat. In other words, if a small planet such as the Moon ($R_{\mathrm{P}} = 1740$ km) was extensively melted when short-lived radioactive isotope such as ^{26}Al is not available (more than ~10 Myr after the formation of the solar system), energy source other than its own gravitational energy must be present. The gravitational energy of the Earth (i.e., a giant impact to the Earth) is an obvious possibility to have caused extensive melting during the formation of the Moon.

The temperature given by eqn [15] is the maximum temperature for two reasons. First, in a realistic case, some of the energy is lost by radiation that reduces the temperature (e.g., Hanks and Anderson, 1969). It is understood that in the later stage of planetary accretion, the degree of radiative energy loss is small because of the short timescale of mass addition (due to the collisions of large bodies) and the blanketing effects of debris and/or the atmosphere (e.g., Hayashi et al., 1979; Kaula, 1979; Safronov, 1978). Second, some of the energy will be stored as an internal energy and only a fraction of energy changes to heat. Using a thermodynamic model of shock compression, one can analyze the degree of heating upon a collision. The heating associated with a collision is caused by the adiabatic heating and the entropy production due to compression (shock compression is not isoentropic) and is given by

$$\mathrm{d}T = -\frac{T\gamma}{V}\mathrm{d}V + \frac{1}{2C_v}[(P - P_{\mathrm{o}})\mathrm{d}V + (V_{\mathrm{o}} - V)\mathrm{d}P] \qquad [16]$$

where γ is the Grüneisen parameter, C_v is the specific heat, V is the volume at pressure P, and V_{o} is the volume at pressure P_{o} (Carter et al., 1975). The first term represents the adiabatic heating and the second term corresponds to the heating due to the excess entropy produced by shock deformation. Both of these terms are sensitive to the material properties such as the bulk modulus and the Grüneisen parameter (including its volume dependence) and hence the temperature increase due to a collision depends strongly on the compressional properties of materials (**Figure 26**) (Karato, 2014a). Because the compressional properties are markedly different between liquids and solids (e.g., Jing and Karato, 2011), the consequence of a giant impact will be different if the surface of a planet that is hit has a surface magma ocean or not. Sensitivity of collisional heating on material properties likely affects the composition of materials ejected by a giant impact. In examining the thermal and chemical aspects of a giant impact, the influence of compressional properties (equation of state) of materials involved must be examined carefully (for a discussion of shock heating, see also Ahrens and O'Keefe, 1972; Stevenson, 1987).

Some of the ejected materials by a giant impact is trapped by the gravity of a mother planet (e.g., the Earth) and is confined to a relatively small space (less than 5–10 RE (the Earth's radius) in the case of the Moon formation; Canup, 2004a). As a result, the pressure of the gas (+some condensed matter) is high in comparison with the pressure in the solar nebula, and the condensation in such a case occurs as gas→liquid rather than

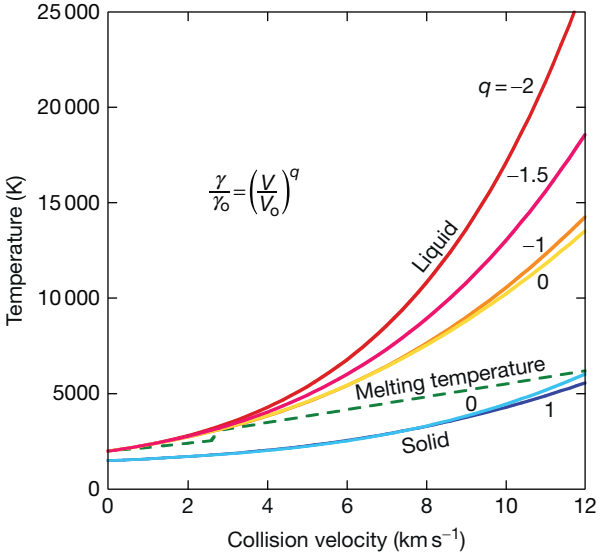


Figure 26 Temperature increase due to a collision as a function of collision velocity (from Karato, 2014a). The collision of a liquid (a planet covered with a magma ocean (proto-Earth)) and a solid at a planar interface is considered. Plausible collision velocity is the escape velocity, ~10–11 km s^{-1} (for the collision to the proto-Earth). A larger degree of heating in liquids than solids is caused by the larger degree of compression for liquids than solids (due to the difference in the bulk modulus) and by the difference in the volume dependence of the Grüneisen parameter $\left(\gamma = \left(\frac{\partial \log T}{\partial \log \rho}\right)_{\mathrm{ad}}\right)$ that represents the efficiency of adiabatic heating (the Grüneisen parameter in silicate melts increases with compression, whereas the Grüneisen parameter in silicate minerals decreases with compression). As a consequence, vaporization due to a collision will occur mostly from the preexisting magma ocean if the proto-Earth was covered with a magma ocean. Heating in the solid planet unlikely exceeds the melting temperature.

gas → solid (see Figure 3). In these cases, a substantial amount of volatile elements can be dissolved in the condensed matter (liquid). If accretion is fast in comparison with the cooling timescale of the disk, then a substantial fraction of volatiles dissolved in the liquid phase could be preserved in the satellite (the Moon) (Karato, 2013a).

9.05.5.2.2 Formation of the atmosphere and the ocean

When the mass of a growing planet is large, it can acquire some gas from the surrounding nebula. However, geochemical observations, particularly those on heavy rare gases (e.g., Ne, Ar, and Xe), show that the amount of this primary gas (the solar gas) trapped by a terrestrial planet is small (e.g., Marty, 2012; Ozima and Igarashi, 1989). The majority of the volatile elements on a terrestrial planet are there as a consequence of outgassing from the planetary interior (e.g., Rubey, 1951).

Abe and Matsui developed a model of formation of the atmosphere and the ocean in the late stage of planetary formation (Abe and Matsui, 1985, 1988; Matsui and Abe, 1986a,b). They emphasized the role of impact-induced degassing and the chemical equilibrium at the interface between the growing magma ocean and the atmosphere. They found that the amount of water contained in the primitive atmosphere formed by the degassing during the formation of the Earth is $\sim 1.2 \times 10^{21}$ kg, which agrees well with the current ocean mass on Earth (1.4×10^{21} kg). They also showed that the predicted ocean mass is only weakly dependent on the amount of water contained in the source materials due to the large capacity of magma ocean to buffer the water content of the atmosphere. The near agreement of the predicted value of the ocean mass with the actual ocean mass on Earth is remarkable, and they suggested that water contained in the impact-induced atmosphere might be the source of the ocean water. If this model were correct, the Earth's interior would be essentially dry.

However, this model has some limitations, and it is not clear if this model explains the amount of volatiles in a planet. First, in their model, volatile elements dissolved in the magma ocean were not considered as a source of volatiles in the atmosphere and ocean. Because the solubility of volatile elements (such as hydrogen and carbon) in the magma is large (Figure 25), the magma ocean can contain a large amount of volatiles, and neglecting the volatile elements dissolved in the magma ocean is difficult to justify. Elkins-Tanton (2008) developed a different model for the atmosphere and the ocean formation on a terrestrial planet where the role of volatile elements dissolved in the magma ocean is emphasized. This model contains some details of the consequence of solidification, and similar to the model by the Abe and Matsui, this model predicts a dense primitive atmosphere exceeding ~10 MPa (~100 bars). In this model, the amount of water in the mantle (at the time of completion of magma ocean solidification) is determined by the amount of initial water content in the magma ocean and the solubility of water (hydrogen) in mantle minerals. Also, such a model predicts compositional stratification including the volatile content (e.g., Elkins-Tanton and Grove, 2011; Ohtani, 1985). However, the degree of stratification depends on the details of crystal settlements in a convecting magma ocean (e.g., Solomatov, 2009), and convection after the complete solidification could also destroy the initial stratification (e.g., Sleep, 1988). Hamano et al. (2013) recently investigated the consequence of magma ocean solidification and emphasized the importance of the cooling rate that determines the amount of water retained in a planet after the solidification of the magma ocean.

Second, Genda and Abe (2005) showed that when a large collision occurs on a planet with preexisting magma ocean, almost all the preexisting atmosphere will be blown off. If this happens, then the volatile content in the initial atmosphere would not play an important role in the long-term evolution of the volatile budget.

Regardless of the details, all models predict the presence of a hot initial atmosphere that helps maintain the surface temperature high enough to melt silicates. Such a high temperature lasts only through the period in which the rate of accretion is high and the large energy release occurs in a growing planet. As the rate of accretion (energy input) reduces, the atmosphere will cool down to allow surface ocean to exist if the greenhouse effect of the atmosphere is limited. The time for cooling depends on the energy release from a growing planet and is estimated to be on the order of 1 Myr for the Earth (e.g., Sleep, 2001). This means that the ocean is likely formed in the very early stage of evolution of a terrestrial planet.

9.05.5.2.3 Influence of core formation

Core formation (i.e., Fe–silicate separation) on Earth and other terrestrial planets occurred in the early stage of planetary evolution (<50 Myr; Kleine and Rudge, 2011). Fe–silicate separation occurs efficiently when the surface of a growing planet is covered with a magma ocean (e.g., Karato and Murthy, 1997; Stevenson, 1990). The core formation has important influence on the distribution of volatiles because iron can dissolve a large amount of volatiles at high pressure and temperature conditions (Section 9.05.2.3).

Two issues are critical in estimating how much volatiles might be in the core. First, in order for a substantial amount of volatiles to be dissolved in the core, the core formation must occur in a growing planet that already has acquired some volatiles. If the volatile acquisition were late as suggested by Albaréde (2009), then not much volatiles would be in the core. Second, even if the core formation occurred after a substantial amount of volatiles is already acquired in a growing planet, the degree to which separating iron captures the volatiles depends on the processes of core formation. Wood and Halliday (2010) presented evidence that a substantial amount of volatiles was already present when the core–mantle separation occurred on Earth. On the other hand, Karato and Murthy (1997) suggested that much of the iron–silicate separation occurs by merging of preexisting cores without chemical equilibrium. In such a case, the amount of volatiles in the core will be small. Dissolved volatiles will affect various physical properties of the core including density, seismic wave velocity, and electrical (and thermal) conductivity (e.g., Shibazaki et al., 2012).

Although currently no constraints are available for the water (and other volatile elements) concentration in the core, it is likely that the core has less than the solubility limits of these volatile elements. This is because the volatile solubility increases with pressure and the core–mantle equilibrium during core formation is likely established at the pressure lower than the current core–mantle boundary pressure (e.g., Karato and Murthy, 1997; Rubie et al., 2007). This means that even if

the core could contain a large amount of volatiles (e.g., water; Figure 7), it likely serves as a *sink* for the mantle water rather than the *source* for the mantle water as far as the degree of undersaturation is larger in the core than in the mantle.

9.05.6 Evolution of the Ocean and the Atmosphere

9.05.6.1 Observational Constraints

Vigorous degassing during the late stage of planetary accretion is a natural consequence of the oligarchic growth of planets. This suggests the formation of the atmosphere and the ocean on a terrestrial planet in the early stage of its evolution (the late stage of planetary accretion). There are a number of observations supporting this view. For example, Hamano and Ozima (1978) used $^{40}Ar/^{36}Ar$ to suggest massive degassing in the early history of the Earth. (The primordial gas does not have much ^{40}Ar, and most of ^{40}Ar comes from the radioactive decay of 40 K in the Earth's interior. Therefore, the abundance of ^{40}Ar (relative to ^{36}Ar) provides a measure of degassing from the Earth's interior.) Using the similar observations, Sarda et al. (1985) estimated the effective time of atmospheric formation of ~4.4 Gyr BP. Allègre et al. (1986/1987) also reported geochemical evidence (the absence of the anomalies in ^{129}Xe) suggesting that the source regions of OIB (presumably the deep mantle) were not as much degassed as the source region of MORB (the asthenosphere) in the early stage of the Earth's evolution.

As with the ocean, using the evidence of low-temperature (~1000 K) crystallization of ~4.3 Gyr old zircon, Watson and Harrison (2005) presented evidence for the presence of the surface ocean in the very early stage of the Earth's evolution. Similarly, based on the oxygen isotopic observation of ~4.3 Gyr old zircon, Mojzsis et al. (2001) presented more indirect inference to reach the same conclusion. Considering the timescale of cooling of hot protoatmosphere of ~1 Myr (Sleep, 2001) and the lifetime of the magma ocean of ~10–100 Myr (e.g., Abe, 1997; Solomatov, 2009), such a short timing for the appearance of the surface ocean is not surprising although the spatial extent of an early ocean is unknown.

After the formation of the primitive atmosphere and ocean in less than the first few 100 million years, a slow evolution of both the atmosphere and the ocean followed for several billions of years. Based on the inferred rate of chemical reactions between volatiles and rocks, Rubey (1951) presented a strong case that the ocean and atmosphere of the Earth must have been formed through the processes involving long-term degassing from the solid Earth, and hence, the ocean mass is maintained by the dynamic balance of these processes. There is a range of observations to suggest the change in the chemistry of both the atmosphere and the ocean during this period (e.g., Berner, 1994; Holland, 2003; Holland et al., 1986; Kasting, 1993), implying the presence of active chemical reactions between the atmosphere–ocean and the solid Earth.

If extensive chemical reactions occur between the atmosphere–ocean and the solid Earth during this period, how did the mass of the ocean and atmosphere change? There are no well-established methods to estimate the history of the atmospheric mass directly. ($^{40}Ar/^{36}Ar$ ratio provides indirect evidence for the evolution of the atmosphere.) Som et al. (2012) used the size of raindrop imprints to infer the air pressure at the end of the Archean period (~2.7 Gyr ago) to conclude that the air pressure on the Earth's surface, that is, the atmospheric mass, at ~2.7 Gyr ago, was not much larger (less than a factor of two difference) than it is today. The D/H ratio also provides some clue as to the extent of atmospheric loss because H is preferentially removed from the atmosphere than D due to a large difference in mass. For instance, the observed much higher D/H for the atmosphere of Venus is considered to be the evidence for a large degree of hydrogen loss (Donahue et al., 1982; Grinspoon, 1987; Kasting and Pollack, 1983).

We have better constraints on the history of the ocean mass. The history of the ocean mass could be inferred from the height of the continental shoreline ('continental freeboard'). If the areas of continent and ocean remain the same and if the average age of the ocean is also constant, then by measuring the height of the shoreline, one can infer the volume (mass) of the ocean. However, the validity of both of these assumptions is unclear, and in fact, some authors used the observed variation of continental freeboard to infer the growth rate of continents using a model of variation of average age of the ocean with time and assuming the ocean mass remained constant (e.g., Reymer and Schubert, 1984).

To understand the reasons for the different conclusions from the same observations (sea-level changes), let us first write an equation among the ocean mass, the relative sea level, and the fraction of continents (Figure 27) (The sea-level equation must include the influence of the ice sheet melting. However, the timescale of climate change (ice age) is much shorter (~1 Myr) than the timescale of ocean volume change that we consider here, so the influence of ice sheet melting is averaged out in this equation except for the period of snowball Earth):

$$[h(t) + X(t) - \bar{z}(t)][1 - f(t)] \cdot S = V(t) \qquad [17]$$

where $h(t)$ is the sea level at the continental margin (continental 'freeboard'), $X(t)$ is the dynamic topography of the continent (topography change of the continent caused by the

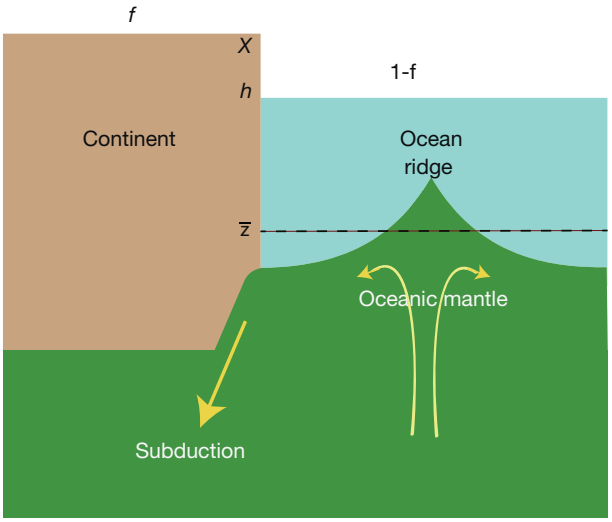


Figure 27 A cartoon showing the definitions of various terms in eqn [17].

convection current beneath (positive upward)), $\bar{z}(t)$ is the average depth of the ocean that depends on the average age of the oceanic mantle, $f(t)$ is the fraction of the surface covered by the continents, S is the surface area of the Earth, and $V(t)$ is the volume of the ocean. I assume that the temperature variation occurs only in the shallow mantle, and therefore, the surface area S remains approximately constant. Equation [17] contains five variables ($h(t)$,$X(t)$, $f(t)$, $\bar{z}(t)$, $V(t)$), and hence, if one were to infer the ocean mass history, $V(t)$, from one observation, $h(t)$, one needs to know the history of the continental area, $f(t)$; the history of the mean ocean depth, $\bar{z}(t)$; and the dynamic topography, $X(t)$.

The continental growth curve, $f(t)$, is controversial but most models suggest that the variation in $f(t)$ is less than ~30% during the whole Proterozoic period (~2.5 Gyr to ~600 Myr BP; e.g., Reymer and Schubert, 1984; Stein and Ben-Avraham, 2007). If one focuses on the Phanerozoic period (~600 Myr ago to the present) where reliable observations on the sea level, $h(t)$, are available, the change in $f(t)$ and its uncertainties are less than 10%, and hence, the influence of this term on the ocean volume is less than ~3%. The change in the mean ocean depth, $\bar{z}(t)$, could be calculated from the decay of heating by radioactivity using the boundary layer theory of internally heated fluid (e.g., Turcotte and Schubert, 1982). For the last ~1 Gyr, the heating rate changed ~15% that translates to a change in $\bar{z}(t)$ of ~4%. Sleep (1992) suggested even a smaller change in the plate velocity with time. However, the short-term fluctuation in the average age of the ocean floor and hence $\bar{z}(t)$ can be larger (e.g., Richards et al., 2000). (In addition to the change in average sea level (eustatic sea-level change), the local effects near the continental margins such as the dynamic topography associated with subduction will also contribute to the observed sea-level changes (e.g., Gurnis, 1993).) In fact, the observed sea-level variations do show large short-term fluctuations (**Figure 28**; e.g., Eriksson et al., 2006; Galer and Mezger, 1998; Hallam, 1992; Parai and Mukhopadhyay, 2012; Wise, 1974). In terms of the absolute rate of sea-level changes, the short-term sea-level change rate is up to ~10 times larger than the long-term sea-level change rate (~10% variation in ~1 Gyr and ~10% variation in ~100 Myr). The influence of dynamic topography on the sea level can be large (several 100 ms) but is sensitive to the viscosity–depth structure and will be discussed later.

The inferred long-term stability of the ocean mass on Earth (near-constant ocean mass for ~1 Gyr or more) is critical for the evolution of life (e.g., Langmuir and Broecker, 2012). However, considering potential causes to destabilize the ocean mass including the climate change and the changes in the degassing and/or regassing rates associated with the changes in plate velocities or the distribution of the continents,

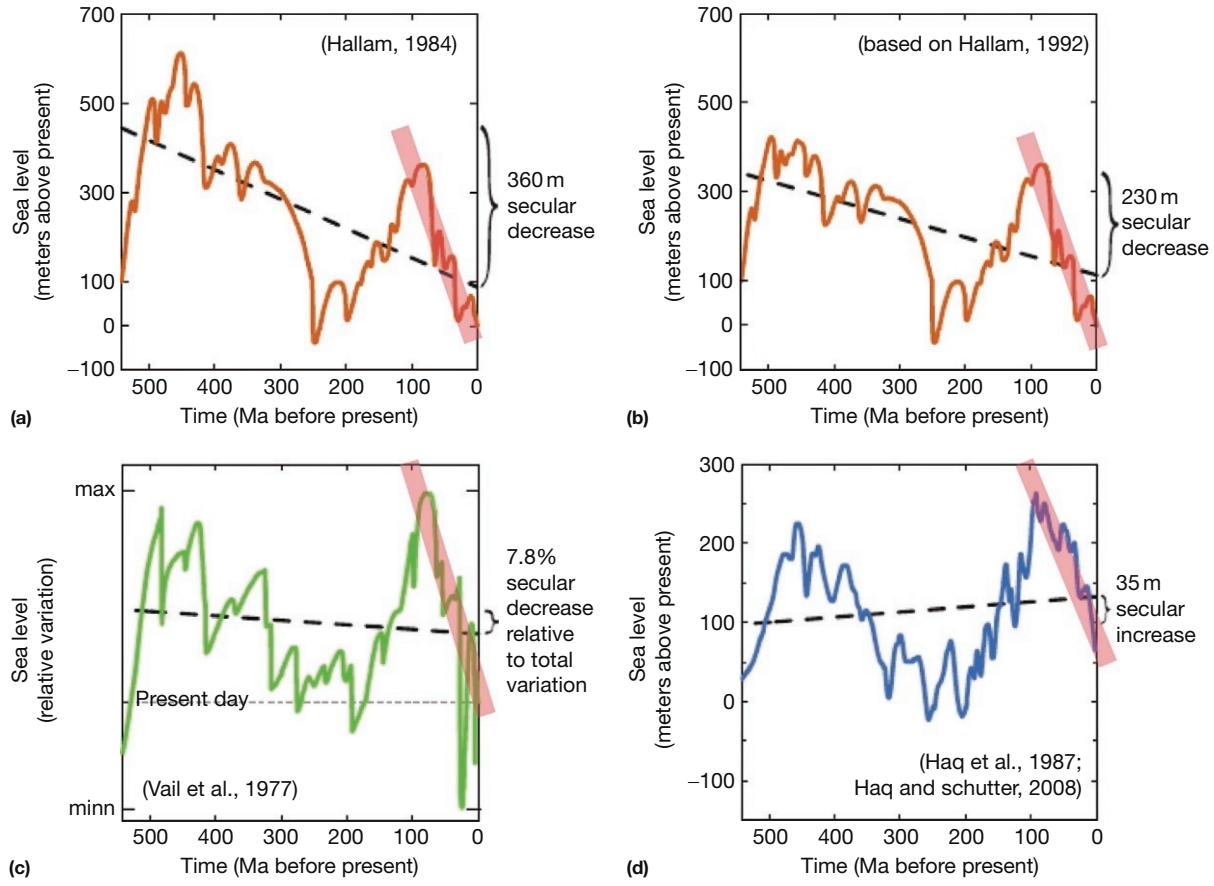


Figure 28 Phanerozoic sea levels compiled by Parai and Mukhopadhyay (2012). In addition to the secular change, large fluctuations are observed. The long-term variation in the sea level is small (less than ~5%/gigayear; black broken lines), whereas short-term variation is large (~5%/100 Myr in the recent ~100 Myr; red thick lines).

the inferred stability of ocean mass is remarkable and requires some explanation. In the following, I will review our current understanding of the long-term stability of ocean mass from two aspects: (i) the stability of atmospheric mass (and temperature) and (ii) the stability of ocean mass through the ocean–solid Earth interactions. However, since the stability of the atmosphere including hydrogen loss from the atmosphere and the conditions for the stable climate has been discussed extensively (e.g., Abe et al., 2011; Hunten and Donahue, 1976; Hunten et al., 1989; Kasting, 1989; Kasting and Catling, 2003; Kasting et al., 1993), I will make only a brief summary on it and focus on the second issue that has not been discussed in detail.

9.05.6.2 Stability of the Atmosphere and the Hydrogen (Water) Loss from the Atmosphere

For the ocean to be present for billions of years, the clement atmosphere must exist for a long time that allows the escape of water (hydrogen) from the atmosphere to be small. For this, a key is the concentration of greenhouse gases such as carbon dioxide and water vapor. Emphasizing the role of carbon dioxide, Walker et al. (1981) proposed that the interaction between the atmosphere and the solid Earth through the formation of carbonates may stabilize the climate: if the concentration of carbon dioxide is high in the atmosphere, then temperature will increase leading to the increased rate of reaction to form carbonate that in turn will reduce the concentration of carbon dioxide in the atmosphere. Conversely, if much of the atmospheric carbon dioxide is stored in carbonate (low carbon dioxide in the atmosphere) and if subduction occurs by plate tectonics, then a large amount of subducted carbonate will increase the outgassing from the magma to increase the amount of atmospheric carbon dioxide. Such a *negative feedback* will stabilize the atmospheric concentration of carbon dioxide and hence the atmospheric temperature (climate). Other factors such as albedo (reflectivity) and the fraction of continents that affects near-surface water content also control the stability of atmospheric temperatures (e.g., Abe et al., 2011; Kasting et al., 1993). Abe et al. (2011) found that for a reasonable range of solar radiation, conditions favoring the presence of oceans could last ~10 Gyr or more.

Important conclusions from these studies are as follows: (i) there are some conditions under which atmosphere with modest temperature to support the surface oceans exists for more than ~10 Gyr, but (ii) under other conditions, substantial atmospheric (or hydrogen) escape occurs. When significant but selective atmospheric loss occurs, there will be an isotopic signature (Hunten et al., 1989; Pepin, 1997; Robert et al., 2003). In particular, the change in the D/H ratio will be large when effective atmospheric loss occurs because hydrogen (H) is easier to escape than deuterium (D) because of the large difference in the mass (Robert et al., 2003). In the following, I will assume that the conditions of the stable presence of atmosphere of modest temperature for the liquid water to exist are met and ask how the fluctuations in degassing and/or regassing rates could affect the ocean mass.

9.05.6.3 Evolution of the Ocean Mass: Models of Global Water (Volatile) Circulation

Under the assumption that the water loss from the atmosphere is negligible, the total amount of water in the ocean and the Earth's interior (mantle) is fixed. Then, the evolution of the ocean mass is the history of exchange of water between the ocean and the solid Earth (mantle), that is,

$$X^{\text{total}} = X^{\text{mantle}} + X^{\text{ocean}} (= \text{constant}) \quad [18]$$

and

$$\frac{dX^{\text{ocean}}}{dt} = \Gamma_D - \Gamma_R \equiv F(X^{\text{ocean}}) \quad [19]$$

where X^{ocean} is the ocean mass, X^{mantle} is the water mass in the mantle (X^{total}: total water mass), Γ_D is the rate of water transport from the mantle to the ocean (*degassing* rate), and Γ_R is the rate of water transport from the ocean to the mantle (*regassing* rate) (Table 4).

Figure 29 illustrates the processes of degassing and regassing. Degassing is caused by partial melting near the surface and most of the volatiles in the mantle rocks are removed to

Table 4 Inferred degassing and regassing rate (global integrated values)[a]

	Values	*References*
Degassing rate[b]	1.5×10^{11} kg year^{-1}	Data compiled by Parai and Mukhopadhyay (2012)
Regassing rate[c]	0.7×10^{12} kg year^{-1}	Jarrard (2003)
	1.1×10^{12} kg year^{-1}	van Keken et al. (2011)
	1.3×10^{12} kg year^{-1}	Hacker (2008)
	$(0.9–1.8) \times 10^{12}$ kg year^{-1}	Rüpke et al. (2004)
	$(2.0–2.9) \times 10^{12}$ kg year^{-1}	Schmidt and Poli (1998)

[a]The estimated values correspond to near recent processes. Degassing rate is likely much faster in the early stage of the Earth's evolution (see **Figure 31**). The current ocean mass is 1.4×10^{21} kg. If this much water were to be transported in 1 Gyr, the rate of transportation would be 1.4×10^{12} kg year^{-1}. With the inferred degassing rate in the early stage of the Earth's evolution (~10^{13} kg year^{-1}), one ocean mass could have been produced in ~100 Myr.

[b]The rate of transportation of water-containing materials to the volcanic regions (e.g., mid-ocean ridges) is well constrained (with the uncertainties of ~50%). In some studies, all of water in these rocks is assumed to degass (e.g., Franck and Bounama, 2001; McGovern and Schubert, 1989; Parai and Mukhopadhyay, 2012), whereas in other studies, only a fraction of it is considered to degass (e.g., Rüpke et al., 2006).

[c]Among various estimates of regassing rate, Jarrard (2003) provided the lowest rate while Schmidt and Poli (1998) showed the highest rate. Jarrard (2003) considered only the sediments and the oceanic crust, whereas Schmidt and Poli (1998) considered the deep oceanic crust and the upper mantle as well including the dense hydrous phases such as phase A. A wide variation in the estimated regassing rate is partly caused by the assumed mineralogy.

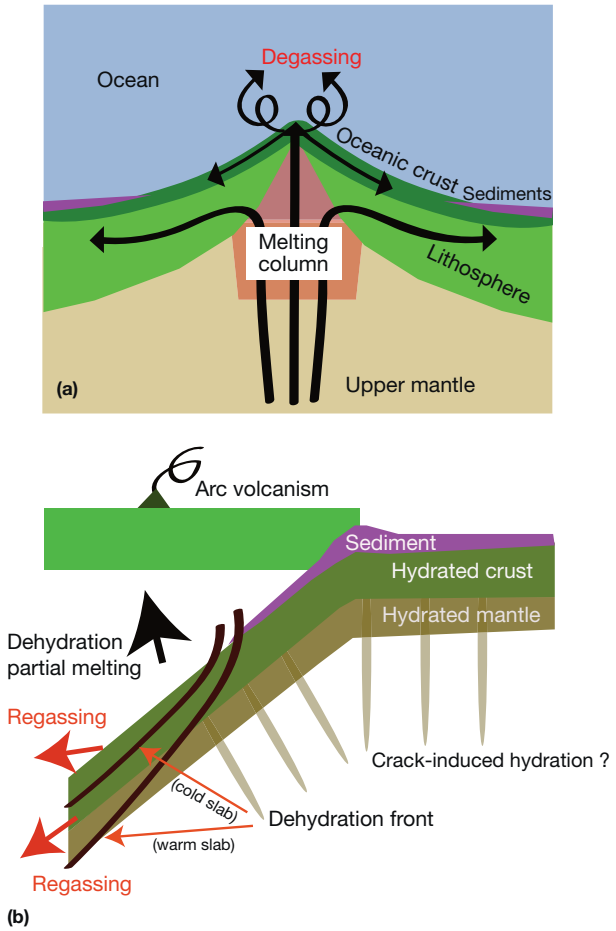


Figure 29 (a) A schematic drawing of processes of degassing at mid-ocean ridges. Water-bearing mantle materials rise nearly adiabatically below volcanoes (e.g., mid-ocean ridges) leading to partial melting. Volatiles go mostly to the melt, and when melt is frozen, some of the dissolved volatiles go to the ocean and atmosphere ('degassing'). (b) A schematic drawing of processes of regassing in the subduction zone. Volatiles such as hydrogen and carbon react with rocks to form hydrous minerals or carbonates. Hydrous minerals (including sediments) and carbonates on the oceanic lithosphere go back to the mantle upon subduction ('regassing'). The rate of regassing depends on the degree of hydration (carbonation) and the degree of dehydration (decarbonation) by heating (**Table 4**).

the molten rock (basalt). When basaltic rocks cool down, most of the volatiles will go to the ocean or to the atmosphere. One uncertainty is the fraction of volatiles in basalt that goes to the ocean (and atmosphere). Some of the volatiles will remain in the deep oceanic crust (gabbro).

Regassing involves more complicated processes than degassing. The regassing occurs at trenches where hydrated oceanic lithosphere carries water (and other volatiles) into the Earth's interior by subduction. The rate of regassing is therefore dependent on the degree of hydration of the oceanic lithosphere and the degree to which hydrated materials survive dehydration by heating after subduction. The degree of hydration of the oceanic lithosphere is controversial. Hydration caused by the hydrothermal activities near mid-ocean ridges is down to ~5 km (e.g., Lowell et al., 1995). In addition to near-ridge hydration, evidence of deep (to ~20 km) hydration by cracking near trenches is reported (Faccenda et al., 2009; Garth and Rietbrock, 2014; Ranero et al., 2003; Van Avendonk et al., 2011) although such processes may not be global. When volatiles subduct deep into the mantle, heating will dehydrate them, and only a fraction of volatiles goes into the deep mantle. However, the degree to which dehydration occurs depends on several parameters including (i) the temperature distribution near the slab/mantle interface that depends on a number of parameters including the age of subducting lithosphere and the degree of mechanical coupling between the slab and the wedge mantle (e.g., Rüpke et al., 2004; van Keken et al., 2011) and (ii) the depth of hydration. Both of which likely depend on various tectonic setting and are expected to vary from one region to another.

Both degassing and regassing rates are proportional to the rate at which materials circulate in the Earth's mantle. Therefore, these models must be combined with a model of evolution of mantle convection (thermal history). In thermal history modeling, the interplay between temperature (and water content)-dependent rheological properties and cooling plays an important role.

There have been several studies to investigate the history of the ocean mass through these approaches (e.g., Crowley et al., 2011; Franck and Bounama, 2001; Korenaga, 2011; McGovern and Schubert, 1989; Rüpke et al., 2006; Wallmann, 2001). In most of these studies, one develops some models for the degassing and regassing rates (Γ_D, Γ_R) and integrates eqn [19] together with the thermal evolution equation.

The ocean mass evolution depends essentially on $\Gamma_D - \Gamma_R$ and the results of such modeling depend entirely on this value. Therefore, subtle variations in either the degassing or the regassing rate (or both) could change the results of such a model drastically. Because a large number of parameters and assumptions are involved in these models, the results of these models must be compared to observations carefully. However, the interpretation of observations is not straightforward. In the following, I will discuss some of the key issues in the study of evolution of the ocean mass.

Most models show initially high degassing rate, $\sim 10^{13}$ kg year^{-1}, that decreases with time as a result of secular cooling (**Figure 30(a)** and **30(b)**). This fast initial degassing rate implies that one ocean mass could have been created in ~100 Myr by degassing from the interior (in addition to water from the condensation of the moist protoatmosphere (Abe and Matsui, 1985) and/or from the solidification of the water-bearing magma ocean (Elkins-Tanton, 2008; Hamano et al., 2013)). One of the large differences among different models is the treatment of the regassing rate. In some models, the dependence of regassing rate on the (average) slab age is not included and the regassing rate and the mantle water content change with time monotonically (e.g., Franck and Bounama, 2001; McGovern and Schubert, 1989, **Figure 30(a)**; Crowley et al., 2011, **Figure 30(c)**). In contrast, Rüpke et al. (2006) included the high sensitivity of the regassing rate on the (average) slab age (temperature) in their model and found that substantial regassing starts only in the later stage of the Earth's history where cold slabs subduct, leading to nonmonotonic variation of mantle water content with time (**Figure 30(b)**). The large influence of regassing parameter on the history of the ocean

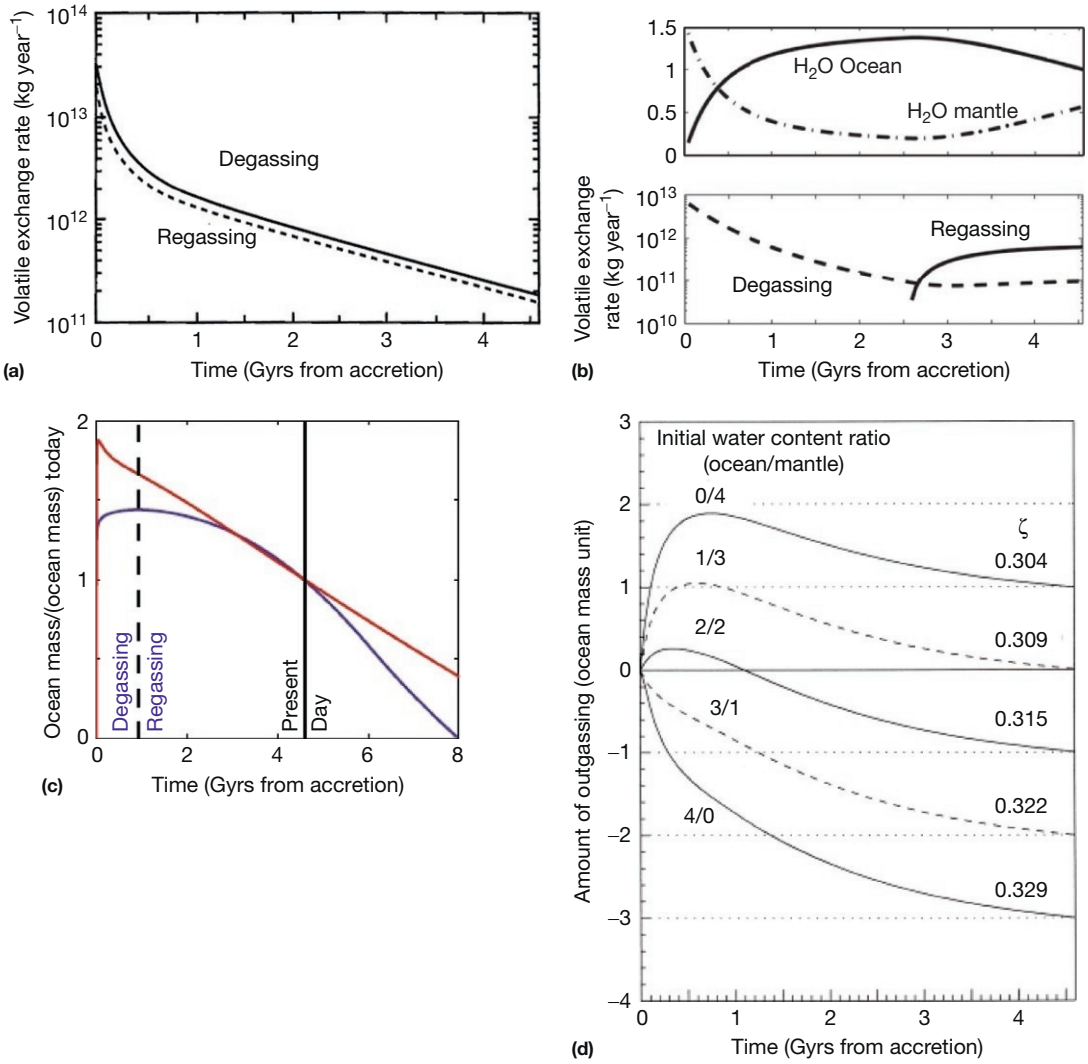


Figure 30 An example of water history (degassing/regassing rate and ocean mass) calculation on Earth. (a) In this model, degassing rate is always higher than regassing rate leading to the increase in the ocean mass (reproduced from McGovern PJ and Schubert G (1989) Thermal evolution of the Earth: Effects of volatile exchange between atmosphere and interior. *Earth and Planetary Science Letters* 96: 27–37). (b) In this model, an efficient regassing occurs only after the Earth has cooled enough. Therefore, the ocean mass increases with time initially but decreases later (reproduced from Rüpke LH, Phipps Morgan J, and Dixon JE (2006) Implications of subduction rehydration for Earth's deep water cycle. In: Jacobsen SD and Lee Svd (eds), *Earth's Deep Water Cycle*, pp. 263–276. Washington, DC: American Geophysical Union). (c) Ocean mass history for water-dependent rheology (blue line) and for water-independent rheology (red line). Regassing efficiency is assumed to be constant (reproduced from Crowley JW, Gérault M, and O'Connell RJ (2011) On the relative role of influence of heat and water transport on planetary dynamics. *Earth and Planetary Science Letters* 310: 380–388). (d) The amount of total outgassed water is plotted against time for various initial ocean water mass/initial mantle water. ζ is a fraction of subducted water that goes into the mantle. This result shows high sensitivity of ocean mass history to the regassing rate (reproduced from Franck S and Bounama C (2001) Global water cycle and Earth's thermal evolution. *Journal of Geodynamics* 32: 231–246).

mass was also shown by Franck and Bounama (2001) (**Figure 30(d)**).

A common approach in these studies is to identify a range of parameter set that explains the 'observed' long-term (~1 Gyr) ocean mass evolution (water circulation between the ocean and the mantle). Therefore, the 'stability' of the ocean mass is a part of the assumptions in these models, and the cause for the stability was never considered. However, the choice (or the interpretation) of 'observational constraints' is not trivial. The issues on the estimated mantle water content are discussed in **Section 9.05.4**. Rüpke et al. (2006) chose a small water content in the mantle (~30% ocean mass) consistent with a model by Dixon et al. (2002), but Franck and Bounama (2001) chose a higher water content (approximately three times ocean mass) based on the discussion by Ringwood (1975). As I discussed in **Section 9.05.4**, both geochemical and geophysical observations favor a high water content in the mantle (approximately three times ocean mass or higher).

Perhaps more important is the interpretation of the short-term variations of the sea level. Given the high sensitivity of the ocean mass evolution to the regassing rate and given the sensitivity of regassing rate on a number of variables (processes),

one may expect a large fluctuation in the ocean mass. Indeed, the reported sea level (continental freeboard) history shows very large short-term (~100 Myr) variations (Figure 28). In most of previous studies, however, the reported short-term sea-level variations were attributed to some 'tectonic' effects (effects due to $\bar{z}(t)$, $X(t)$ in eqn [17]), and they considered that the ocean mass did not change appreciably. For instance, Gurnis (1993) and Worsley et al. (1984) noted a good correlation between the distribution of continents and the sea level versus time curve and argued that the observed fluctuation of the sea level is largely due to the influence of the change in the distribution of continents. In more detail, Gurnis (1993) noted that the dynamic topography near subduction zones plays an important role in the observed sea-level changes and could explain the observed sea-level variation assuming that the viscosity–depth profile in these regions is the same as other regions. However, Billen and Gurnis (2001) showed that the low viscosity–depth structure of the upper mantle in these regions substantially reduces the amplitude of dynamic topography. I conclude that if low viscosity in these regions is included, then it will be difficult to explain the observed large sea-level variation by the dynamic topography. It is likely that at least a part of these short-term sea-level fluctuations is caused by the change in the ocean mass.

Considering these uncertainties, it is worthwhile to discuss some possible mechanisms to 'stabilize' the ocean mass physically. Emphasizing the role of regassing on the ocean mass, Kasting and Holm (1992) proposed that the ocean mass may be stabilized due to the influence of ocean mass (pressure at the bottom of the ocean) on the degree of hydration near the mid-ocean ridges. Hydration of the oceanic lithosphere occurs mostly by the hydrothermal circulation near mid-ocean ridges (e.g., Lister, 1980) and the degree to which hydration occurs is sensitive to the properties of aqueous fluids. Kasting and Holm (1992) noted that the pressure and temperature of hydrothermal fluids at the present time are close to the critical conditions of water where water can carry unusually high amount of heat. Consequently, the depth extent to which hydrothermal circulation penetrates is largest under the current conditions, leading to a negative feedback when the ocean depth is below the current value. However, if the ocean mass exceeds the current level, the ocean mass would increase and will not be stabilized. In addition, even if the degree of hydration is controlled by this mechanism, the regassing rate also depends on the near-surface temperature of subducting plate where substantial dehydration occurs. Therefore, even if the degree of hydration is stabilized, a large fluctuation in the degree of dehydration will cause the large fluctuation in the regassing rate. Alternatively, the ocean mass might be regulated by the negative feedback between ocean mass and the degassing efficiency. The degassing efficiency depends on the pressure at which degassing occurs from the erupted magma. When the ocean mass increases, then pressure increases leading to less efficient degassing that causes the decrease in the ocean mass and vice versa.

One of the major limitations of the most previous models of global water circulation is the assumption that the mantle acts as a single reservoir (Figure 31(a)). In such a case, the reaction of the mantle to a change in regassing rate is direct. In contrast, when there are internal processes to redistribute water, then the reaction of the mantle to the variation in the regassing rate is indirect and, under some cases, reduces the sensitivity of the mantle water content to the change in the regassing rate. One example is the role of midmantle partial melting (Figure 31(b)). Partial melting buffers the water content in the residual mantle (see Figure 12) and hence the sensitivity of the mantle water content to the fluctuation in the regassing rate. The observational support for midmantle partial melting and its consequence on the global water circulation were discussed by Karato (2011). Water-promoted melting could occur also below the '660 km' discontinuity in or near the subduction zone. Partial melting below '660 km' discontinuity is expected to be limited to near the subduction zone. Seismological tests of the presence of partial melting below the '660 km' discontinuity is limited (e.g., Schmandt et al., 2011) compared with a layer above the '410 km' discontinuity (e.g., Bagley et al., 2009; Coutier and Revenaugh, 2007; Tauzin et al., 2010). Internal processes such as midmantle partial melting are a consequence of the presence of a peak in water solubility in the midmantle as shown in Figure 7. If indeed such internal processes as deep partial melting were essential for the long-term presence of the surface ocean, it would imply that the planetary size would need to be large enough to cause such partial melting (pressure in the mantle needs to exceed ~15 GPa). Numerical modeling combined with some observational constraints will be helpful to understand these potentially important processes.

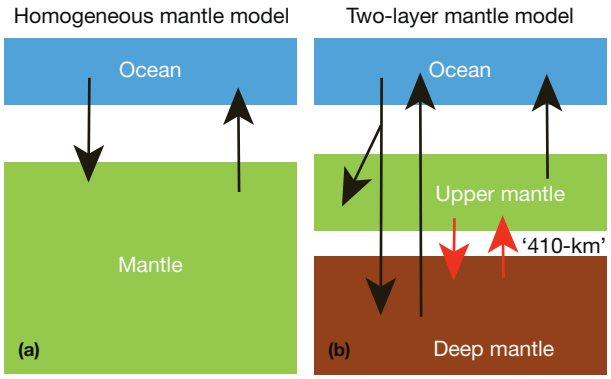


Figure 31 A schematic diagram showing the processes of water circulation in Earth. (a) A single-layer mantle model. (b) A two-layer mantle model where partial melting at '410 km' redistributes water.

9.05.7 Origin of Volatiles on Earth and Other Terrestrial Planets

Where did volatile elements (such as H) on Earth and other terrestrial planets come from? Are the volatile elements acquired by the Earth mostly from materials in or close to the current Earth's orbit (1 AU from the Sun)? Or was there substantial mixing of materials during planetary formation such that the Earth acquired its volatiles mostly from materials away from its current orbit?

The isotopic compositions provide important constraints on the origin of volatiles. Figure 32 summarizes isotopic compositions of two light elements (H and N; from Marty, 2012; Saal et al., 2013). It is seen that comets, meteorites, and

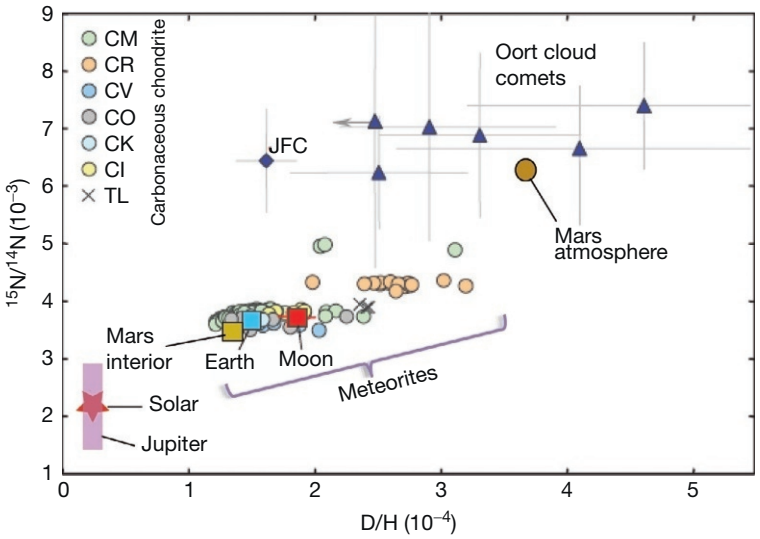


Figure 32 Isotopic compositions of volatile elements H and N of various planetary bodies (after Marty, 2012; Saal et al., 2013). Isotopic compositions of H and N for the Earth, the Moon, and Mars (interior) are similar to the average of meteorites but differ significantly from those of comets. Most of terrestrial planets (the Earth, the Moon, and Mars) are likely formed from materials similar to meteorites. Small contributions from solar component and from comet-like materials are also suggested (e.g., Halliday, 2013; Marty, 2012). CM, CR, CV, CO, CK, CI, and TL represent various types of carbonaceous chondrite. JFC represents Jupiter family comets.

terrestrial planetary bodies (the Earth, the Moon, and Mars) have isotopic compositions distinct from the solar composition, and each terrestrial body has different isotopic compositions. This indicates that terrestrial bodies acquired these light elements whose isotopic compositions were modified from the solar values after the formation of the Sun. Small but clear differences in the isotopic composition between Mars (1.5 AU) and the Earth (1 AU) are noted (see also Pahlevan and Stevenson, 2007), suggesting that the isotopic ratios of these elements strongly depend on the location of these bodies in the solar system.

From such a plot, it is clear that most comet-like materials cannot be a major source of volatiles for the terrestrial planets. (Halliday (2013), however, argued that a contribution from comet-like materials is needed to explain some heavy rare gas (e.g., Kr) observations.) Isotopic signatures of volatile elements of the Earth and other terrestrial planets are close to those of carbonaceous chondrites. However, both Drake and Righter (2002) and Marty (2012) concluded that the contribution from carbonaceous chondrite should be less than ~2%; otherwise, some other geochemical observations cannot be explained (e.g., Os isotopic ratios). Currently, materials similar to carbonaceous chondrites are identified in the outer region of the asteroid belt (~2.5–4.5 AU) (Gradie and Tedesco, 1982). Carbonaceous chondrite-like materials could provide volatiles to the Earth if substantial mixing (due to orbital scattering) occurred in the later stage of planetary formation (~100 Myr after the formation of the Sun) (e.g., Albaréde, 2009; Morbidelli et al., 2000). However, Wood et al. (2010) and Wood and Halliday (2010) presented discussions against the late acquisition of volatiles based on the abundance of siderophile (iron-loving) element abundance pattern and on the different interpretation of the lead isotope data.

I conclude that volatile-rich carbonaceous chondrite-type materials are good candidate from which terrestrial planets acquired volatile elements but that the geochemical observations summarized in the preceding text do not require carbonaceous chondrite-type materials as a source for volatiles. And in fact, there are some geochemical observations suggesting that volatiles were acquired in the early stage of planetary accretion where extensive mixing caused by large orbital scattering is unlikely. Most ordinary chondrites have enough (~0.1 wt%) water (and other volatiles) to make a planet like the Earth with ocean and other volatiles. Based on the Os isotopic observations, Drake (2005) argued against the 'late veneer' origin of volatiles and proposed a possible mechanism to acquire a small amount of volatiles within the snow line through the surface adsorption. By the same token, Marty (2012) showed the presence of the 'solar component' in the Earth's mantle, suggesting that a small amount of the solar gas was trapped in the building block of the Earth. But he also discussed that the volatile (including the rare gas) composition suggests that the volatiles in the terrestrial bodies are made of a mixture of the solar component and the chondritic component.

9.05.8 Summary and Outlook

In this article, I have reviewed the current understanding of the role of water (and other volatile elements) in the evolution of the Earth and other terrestrial planets including the processes of planetary formation and the long-term evolution through global volatile circulation. Studies in such an area require an integration of methodologies and observations in the vast areas such as astronomy, planetary science, geophysics, geochemistry, and materials science (mineral physics). There has been major progress in these areas including (i) the improved understanding of the properties of volatile elements under broad planetary conditions, (ii) the improved

observations to shed some insights into the behavior of volatiles in the planetary environment, and (iii) the development of models where the results of these areas are integrated. However, much needs to be done to answer our major question of why is the Earth so unique in having a modest amount of water on its surface for a geologically long time (billions of years), but not on other terrestrial planets. Let me list some issues that deserve further investigation:

(1) Volatiles in planetary formation

Spectroscopic observations on volatiles on planetary nebulae will help obtain some ideas about the distribution of volatiles in the planetary nebulae under a variety of conditions (e.g., Hrivnak et al., 2000; Miranda et al., 2001).

Previous condensation modeling ignored much of the chemical reactions including volatile elements. Improved modeling based on the experimental data on the equilibrium and kinetic properties (kinetics of chemical reactions) of volatile-bearing materials will be useful.

(2) Water distribution in the deep Earth

Water distribution in Earth has been inferred only for the shallow mantle (upper mantle and transition zone). It is important to infer the water content in the lower mantle and in the core. To do this, experimental studies need to be made to determine the behavior of water (hydrogen) under these conditions, and one must identify and develop methods of remote sensing to infer the water content in these regions.

(3) Volatile distribution in other terrestrial planets

Volatile (water) distribution in Earth (and the Moon) has been inferred using electrical conductivity and tidal dissipation. These properties can be inferred for other planets using remote sensing. It will be important to infer water content (distribution) in Mars, Mercury, and Venus using these techniques.

(4) Modeling volatile circulation and evolution

Only a simplified version of global volatile circulation has been used in most previous studies where the mantle is treated as a single unit. Recent geophysical and geochemical studies summarized here suggest that the Earth's mantle has more complex chemical structure including a layered structure. The nature of water (volatiles) circulation is quite different between a homogeneous mantle model and a layered mantle model. Much effort needs to be focused on modeling in such a realistic system in combination with the improved observations. Particularly important observational constraints include (i) the history of the ocean mass (as discussed in this article) and (ii) the history of the water content in the mantle that might be inferred from the isotopic analysis of ancient MORB.

Acknowledgment

I thank David Evans, Lindy Elkins-Tanton, Bernard Marty, Shigenori Maruyama, Sujoy Mukhopadhyay, Hiroko Nagahara, Masao Nakada, Kazuhito Ozawa, Kaveh Pahlevan, Alberto Saal, and Slava Solomatov for their helpful discussions and Dave Stevenson for inviting me to write this review and for his valuable comments. This work is partially supported by NSF.

References

Abe Y (1997) Thermal and chemical evolution of terrestrial magma ocean. *Physics of Earth and Planetary Interiors* 100: 27–39.

Abe Y, Abe-Ouchi A, Sleep NH, and Zahnle KJ (2011) Habitable zone limits for dry planets. *Astrobiology* 11: 443–460.

Abe Y and Matsui T (1985) Formation of an impact-generated H_2O atmosphere and its implications for the early thermal history of the Earth. *Journal of Geophysical Research* 90: 545–559.

Abe Y and Matsui T (1988) Evolution of an impact-generated H_2O–CO_2 atmosphere and formation of a hot proto-ocean on Earth. *Journal of Atmospheric Sciences* 45: 3081–3101.

Abe Y, Ohtani E, Okuchi T, Righter K, and Drake MJ (2000) Water in the early Earth. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 413–433. Tucson: University of Arizona Press.

Ahrens TJ and O'Keefe JD (1972) Shock melting and vaporization of lunar rocks and minerals. *The Moon* 4: 214–249.

Aizawa Y, Barnhoorn A, Faul UH, Fitz Gerald JD, Jackson I, and Kovács I (2008) Seismic properties of Anita Bay dunite: An exploratory study of the influence of water. *Journal of Petrology* 49: 841–855.

Albaréde F (2009) Volatile accretion history of the terrestrial planets and dynamic implications. *Nature* 461: 1227–1233.

Allègre CJ, Hofmann AW, and O'Nions K (1996) The argon constraints on mantle structure. *Geophysical Research Letters* 23: 3555–3557.

Allègre CJ, Staudacher T, and Sarda P (1986/1987) Rare gas systematics: Formation of the atmosphere, evolution and structure of the Earth's mantle. *Earth and Planetary Science Letters* 81: 127–150.

Baba K, Chave AD, Evans RL, Hirth G, and Randall L (2006) Mantle dynamics beneath the East Pacific Rise at 17°S: Insights from the Mantle Electromagnetic and Tomography (MELT) experiments. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2004JB003598.

Baba K, Utada H, Goto T, Kasaya T, Shimizu H, and Tada N (2010) Electrical conductivity imaging of the Philippine Sea upper mantle using seafloor magnetotelluric data. *Physics of the Earth and Planetary Interiors* 183: 44–62.

Bagley B, Coutier AM, and Revenaugh J (2009) Melting in the deep upper mantle oceanward of the Honshu slab. *Physics of the Earth and Planetary Interiors* 175: 137–144.

Banks RJ (1969) Geomagnetic variations and the electrical conductivity of the upper mantle. *Geophysical Journal of Royal Astronomical Society* 17: 457–487.

Baxter EF (2010) Diffusion of noble gases in minerals. *Reviews in Mineralogy and Geochemistry* 72: 509–557.

Bell DR and Rossman GR (1992) Water in Earth's mantle: The role of nominally anhydrous minerals. *Science* 255: 1391–1397.

Bercovici D and Karato S (2003) Whole mantle convection and transition-zone water filter. *Nature* 425: 39–44.

Berner RA (1994) 3GEOCARB II: A revised model for atmospheric CO_2 over Phanerozoic time. *American Journal of Science* 294: 56–91.

Billen MI and Gurnis M (2001) A low viscosity wedge in subduction zones. *Earth and Planetary Science Letters* 193: 227–236.

Blacic JD (1972) Effects of water in the experimental deformation of olivine. In: Heard HC, Borg IY, Carter NL, and Raleigh CB (eds.) *Flow and Fracture of Rocks*, pp. 109–115. Washington, DC: American Geophysical Union.

Bolfan-Casanova N (2005) Water in the Earth's mantle. *Mineralogical Magazine* 69: 229–257.

Bowen NL (1928) *The Evolution of Igneous Rocks*. Princeton: Princeton University Press, 334 pp.

Burbidge EM, Burbidge GR, Fowler WA, and Hoyle F (1957) Synthesis of the elements in stars. *Review of Modern Physics* 4: 547–650.

Cadek O and Fleitout L (2003) Effect of lateral viscosity variation in the top 300 km on the geoid and dynamic topography. *Geophysical Journal International* 152: 566–580.

Cameron AGW (1988) Origin of the solar system. *Annual Review of Astronomy and Astrophysics* 26: 441–472.

Cameron AGW and Pine MR (1973) Numerical models of the primitive solar nebula. *Icarus* 18: 377–406.

Campbell AJ, Humayun M, and Weisberg MK (2002) Siderophile element constraints on the formation of metal in the metal-rich chondrites Bencubbin, Weatherford, and Gujba. *Geochimica et Cosmochimica Acta* 66: 647–660.

Canup RM (2004a) Dynamics of lunar formation. *Annual Review of Astronomy and Astrophysics* 42: 441–475.

Canup RM (2004b) Simulation of a late lunar-forming impact. *Icarus* 168: 433–456.

Carmichael ISE, Turner FJ, and Verhoogen J (1974) *Igneous Petrology*, p. 739. New York: McGraw-Hill.

Carter WJ, Fritz JN, Marsh SP, and McQueen RG (1975) Hugoniot equation of state of the lanthanides. *Journal of Physics and Chemistry of Solids* 36: 741–752.

Cassen P (1996) Models for the fractionation of moderately volatile elements in the solar system. *Meteoritics and Planetary Sciences* 31: 793–806.

Cemič L, Will G, and Hinze E (1980) Electrical conductivity measurements on olivines Mg_2SiO_4–Fe_2SiO_4 under defined thermodynamic conditions. *Physics and Chemistry of Minerals* 6: 95–107.

Chakraborty S (2010) Diffusion coefficients in olivine, wadsleyite and ringwoodite. *Reviews in Mineralogy and Geochemistry* 72: 603–639.

Chen S, Hiraga T, and Kohlstedt DL (2006) Water weakening of clinopyroxene in the dislocation creep regime. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JB003885.

Chen J, Inoue T, Yurimoto H, and Weidner DJ (2002) Effect of water on olivine-wadsleyite phase boundary in the (Mg, $Fe)_2SiO_4$ system. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2001GRL014429.

Cherniak DJ and Dimanov A (2010) Diffusion in pyroxene, mica and amphibole. *Review in Mineralogy and Geochemistry* 72: 641–690.

Cooper RF and Kohlstedt DL (1982) Interfacial energies in the olivine–basalt system. In: Akimoto S and Manghnani MH (eds.) *High Pressure Research in Geophysics*, pp. 217–228. Tokyo: Venter for Academic Publication.

Costa F and Chakraborty S (2008) The effect of water on Si and O diffusion rates in olivine and implications for transport properties and processes in the upper mantle. *Physics of the Earth and Planetary Interiors* 166: 11–29.

Coutier AM and Revenaugh J (2006) A water-rich transition zone beneath the eastern United States and Gulf of Mexico from multiple ScS reverberations. In: Jacobsen SD and van der Lee S (eds.) *Earth's Deep Water Cycle*, pp. 181–193. Washington, DC: American Geophysical Union.

Coutier AM and Revenaugh J (2007) Deep upper-mantle melting beneath the Tasman and Coral Seas detected with multiple ScS reverberations. *Earth and Planetary Science Letters* 259: 66–76.

Crowley JW, Gérault M, and O'Connell RJ (2011) On the relative role of influence of heat and water transport on planetary dynamics. *Earth and Planetary Science Letters* 310: 380–388.

Dai L and Karato S (2009a) Electrical conductivity of orthopyroxene: Implications for the water content of the asthenosphere. *Proceedings of the Japan Academy* 85: 466–475.

Dai L and Karato S (2009b) Electrical conductivity of pyrope-rich garnet at high temperature and pressure. *Physics of the Earth and Planetary Interiors* 176: 83–88.

Dai L and Karato S (2009c) Electrical conductivity of wadsleyite under high pressures and temperatures. *Earth and Planetary Science Letters* 287: 277–283.

Dai L and Karato S (2014) Highly anisotropic electrical conductivity of the asthenosphere caused by hydrogen diffusion in olivine. *Earth and Planetary Science Letters*, in press.

Dalton CA, Ekström G, and Dziewonski AM (2009) Global seismological shear velocity and attenuation: A comparison with experimental observations. *Earth and Planetary Science Letters* 284: 65–75.

Dasgupta R and Hirschmann MM (2006) Melting in the Earth's deep mantle caused by carbon dioxide. *Nature* 440: 659–662.

Dasgupta R and Hirschmann MM (2010) The deep carbon cycle and melting in Earth's interior. *Earth and Planetary Science Letters* 298: 1–13.

de Pater I and Lissauer JJ (2010) *Planetary Sciences*, 2nd edn., p. 647. Cambridge: Cambridge University Press.

DeMeo FE and Carry B (2014) Solar system evolution from compositional mapping of the asteroid belt. *Nature* 505: 629–634.

Demouchy S, Jacobsen SD, Gaillard F, and Stern CR (2006) Rapid magma ascent recorded by water diffusion profiles in mantle olivine. *Geology* 34: 429–432.

Deuss A (2009) Global observations of mantle discontinuities using SS and PP precursors. *Surveys in Geophysics* 30: 301–326.

Dixon JE, Dixon TH, Bell DR, and Malservisi R (2004) Lateral variation in upper mantle viscosity: Role of water. *Earth and Planetary Science Letters* 222: 451–467.

Dixon JE, Leist L, Langmuir J, and Schiling JG (2002) Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* 420: 385–389.

Dobson D (2003) Oxygen ionic conduction in $MgSiO_3$ perovskite. *Physics of the Earth and Planetary Interiors* 139: 55–64.

Dobson DP and Brodholt JP (2000) The electrical conductivity of the lower mantle phase magnesiowüstite at high temperatures and pressures. *Journal of Geophysical Research* 105: 531–538.

Donahue TM, Hoffman JH, Hodges RR Jr., and Watson AJ (1982) Venus was wet: A measurement of the ratio of deuterium to hydrogen. *Science* 216: 630–633.

Drake MJ (2005) Origin of water in the terrestrial planets. *Meteoritics and Planetary Sciences* 40: 519–527.

Drake MJ and Righter K (2002) Determining the composition of the Earth. *Nature* 416: 39–44.

Duba A and Shankland TJ (1982) Free carbon and electrical conductivity in the Earth's mantle. *Geophysical Research Letters* 9: 1271–1274.

Dziewonski AM (2000) Global seismic tomography: Past, present and future. In: Boschi E, Ekström G, and Morelli A (eds.) *Problems in Geophysics for the Next Millennium*, pp. 289–349. Roma: Editorice Compositori.

Ebel DS and Grossman L (2000) Condensation in dust-enriched systems. *Geochimica et Cosmochimica Acta* 64: 339–366.

Ekström G and Dziewonski AM (1998) The unique anisotropy of the Pacific upper mantle. *Nature* 394: 168–172.

Elkins-Tanton LT (2008) Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth and Planetary Science Letters* 271: 181–191.

Elkins-Tanton L and Grove TL (2011) Water (hydrogen) in the lunar mantle: Results from petrology and magma ocean modeling. *Earth and Planetary Science Letters* 307: 173–179.

Eriksson PG, Mazumder R, Catuneau O, Bumby AJ, and Ilondo BO (2006) Precambrian continental freeboard and geologic evolution: A time perspective. *Earth Science Reviews* 79: 165–204.

Evans RL, Jones AG, Garcia X, et al. (2011) Electrical lithosphere beneath the Kaapvaal craton, southern Africa. *Journal of Geophysical Research* 116. http//dx.doi.org/10.1029/2010JB007883.

Faccenda M, Gerya TV, and Burlini L (2009) Deep slab hydration induced by bending-related variations in tectonic pressure. *Nature Geoscience* 2: 790–793.

Farver JR (2010) Oxygen and hydrogen diffusion in minerals. *Reviews in Mineralogy and Geochemistry* 72: 447–507.

Fegley B (2000) Kinetics of gas-grain reactions in the solar nebula. *Space Science Reviews* 92: 177–200.

Fei H, Wiedenbeck M, Yamazaki D, and Katsura T (2013) Small effect of water on upper-mantle rheology based on silicon self-diffusion coefficients. *Nature* 498: 213–216.

Fiquet G, Guyot F, Kunz M, Matas J, Andrault D, and Hanfland M (2002) Structural refinements of magnesite at very high pressure. *American Mineralogist* 87: 1261–1265.

Franck S and Bounama C (2001) Global water cycle and Earth's thermal evolution. *Journal of Geodynamics* 32: 231–246.

Frost DJ and Dolejš D (2007) Experimental determination of the effect of H_2O on the 410-km seismic discontinuity. *Earth and Planetary Science Letters* 256: 182–195.

Fukai Y and Suzuki T (1986) Iron–water interaction under high pressure and its implications in the evolution of the Earth. *Journal of Geophysical Research* 91: 9222–9230.

Gaillard F, Malki M, Iacono-Maziano G, Pichavant M, and Scaillet B (2008) Carbonatite melts and electrical conductivity in the asthenosphere. *Science* 322: 1363–1365.

Galer SJG and Mezger K (1998) Metamorphism, denudation and sea level in the Archean and cooling of the Earth. *Precambrian Research* 92: 389–412.

Galy A, Young ED, Ash RD, and O'Nions RK (2000) The formation of chondrules at high gas pressures in the solar nebula. *Science* 290: 1751–1753.

Gando A, et al. (2011) Partial radiogenic heat model for Earth revealed by geoneutrino measurements. *Nature Geoscience* 4: 647–651.

Ganguly J and Bose K (1995) Kinetics of formation of hydrous phyllosilicates in the solar nebula. In: *Proceedings of the Lunar Science Conference, vol. XXVI,* pp. 441–442.

Gao W, Matzel E, and Grand SP (2006) Upper mantle seismic structure beneath eastern Mexico determined from P and S waveform inversion and its implications. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2006JB004304.

Garcia RF, Gagnepain-Beyneix J, Chevrot S, and Lognonné P (2011) Very preliminary reference Moon model. *Physics of the Earth and Planetary Interiors* 188: 96–113.

Garth T and Rietbrock A (2014) Order of magnitude increase in subducted H_2O due to hydrated normal faults within the Wadati-Benioff zone. *Geology* 42: 207–210.

Genda H and Abe Y (2005) Enhanced atmospheric loss on protoplanets at the giant impact phase in the presence of oceans. *Nature* 433: 842–844.

Gradie J and Tedesco E (1982) Compositional structure of the asteroid belt. *Science* 216: 1405–1407.

Griggs DT and Blacic JD (1965) Quartz: Anomalous weakness of synthetic crystals. *Science* 147: 292–295.

Grinspoon DH (1987) Was Venus wet? Deuterium reconsidered. *Science* 238: 1702–1704.

Grossman L (1972) Condensation in the primitive solar nebula. *Geochemica and Cosmochemica Acta* 36: 597–619.

Grossman L and Larimer JW (1974) Early chemical history of the solar system. *Review of Geophysics and Space Physics* 12: 71–101.
Gurnis M (1993) Phanerozoic marine inundation of continents driven by dynamic topography above subducting slabs. *Nature* 364: 589–593.
Hacker BR (2008) H_2O subduction beyond arcs. *Geochemistry, Geophysics, Geosystems* 9. http://dx.doi.org/10.1029/2007GC001707.
Haggerty SE and Sautter V (1990) Ultra deep (>300 km) ultramafic, upper mantle xenoliths. *Science* 248: 993–996.
Hallam A (1992) *Phanerozoic Sea-Level Changes*, p. 266. New York: Columbia University Press.
Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.
Hamano K, Abe Y, and Genda H (2013) Emergence of two types of terrestrial planet on solidification of magma ocean. *Nature* 497: 607–610.
Hamano Y and Ozima M (1978) Earth-atmosphere evolution model based on Ar isotopic data. In: Alexander JEC and Ozima M (eds.) *Terrestrial Rare Gases*, pp. 155–171. Tokyo: Japan Scientific Society Press.
Hanks TC and Anderson DL (1969) The early thermal history of the Earth. *Physics of the Earth and Planetary Interiors* 2: 19–29.
Harte B (2010) Diamond formation in the deep mantle: The record of mineral inclusions and their distribution in relation to mantle dehydration zones. *Mineralogical Magazine* 74: 189–215.
Hauri EH, Gaetani GA, and Green TH (2006) Partitioning of water during melting of the Earth's upper mantle at H_2O-undersaturated conditions. *Earth and Planetary Science Letters* 248: 715–734.
Hauri EH, Weinreich T, Saal AE, Rutherford MC, and Van Orman JA (2011) High pre-eruptive water contents preserved in lunar melt inclusions. *Science* 333: 213–215.
Hayashi C (1961) Stellar evolution in early phases of gravitational contraction. *Publications of the Astronomical Society of Japan* 13: 450–458.
Hayashi C (1981) Structure of the solar nebula, growth and decay of magnetic fields and effects of magnetic and turbulent viscosities on the nebula. *Progress of Theoretical Physics Supplement* 70: 35–53.
Hayashi C, Nakazawa K, and Mizuno H (1979) Earth's melting due to the blanketing effect of the primordial dense atmosphere. *Earth and Planetary Science Letters* 43: 22–28.
Hayashi C, Nakazawa K, and Nakagawa Y (1985) Formation of the solar system. In: Black DC and Matthews MS (eds.) *Protoplanets and Planets II*, pp. 1100–1153. University of Arizona Press.
Hirschmann MM (2006) Water, melting, and the deep Earth H_2O cycle. *Annual Review of Earth and Planetary Sciences* 34: 629–653.
Hirschmann MM (2010) Partial melt in the oceanic low velocity zone. *Physics of the Earth and Planetary Interiors* 179: 60–71.
Hirschmann MM and Dasgupta R (2009) The H/C ratio of Earth's near-surface and deep reservoirs, and consequences for deep Earth volatile cycles. *Chemical Geology* 262: 4–16.
Hofmann AW (1997) Mantle geochemistry: The message from oceanic volcanism. *Nature* 385: 219–228.
Holland HD (2003) The geologic history of seawater. In: Turekian KK and Holland HD (eds.) *Treatise on Geochemistry*, pp. 1–46. Amsterdam: Elsevier.
Holland HD, Lazar B, and McGaffrey M (1986) Evolution of the atmosphere and oceans. *Nature* 320: 27–33.
Hood LL, Herbert F, and Sonett CP (1982) The deep lunar electrical conductivity profile: Structural and thermal inferences. *Journal of Geophysical Research* 87: 5311–5326.
Howard AW, Marcy GW, Johnson JA, et al. (2010) The occurrence and mass distribution of close-in super-Earths, Neptunes, and Jupiters. *Science* 330: 653–655.
Hrivnak BJ, Volk K, and Kwok S (2000) 2–45 micron infrared spectroscopy of carbon-rich proto-planetary nebulae. *The Astrophysical Journal* 535: 275–292.
Huang X, Xu Y, and Karato S (2005) Water content of the mantle transition zone from the electrical conductivity of wadsleyite and ringwoodite. *Nature* 434: 746–749.
Huebner JS, Wiggins LB, and Duba AG (1979) Electrical conductivity of pyroxene which contains trivalent cations: Laboratory measurements and the lunar temperature profiles. *Journal of Geophysical Research* 84: 4652–4656.
Hui H, Peslier AH, Zhang Y, and Neal CR (2013) Water in lunar anorthosites and evidence for a wet early Moon. *Nature Geoscience* 6: 177–180.
Hunten DM and Donahue TM (1976) Hydrogen loss from the terrestrial planets. *Annual Review of Earth and Planetary Sciences* 4: 265–292.
Hunten DM, Donahue TM, Walker JCG, and Kasting JF (1989) Escape of atmospheres and loss of water. In: Atreya SK, Pollack JB, and Matthews MS (eds.) *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 386–422. Tuscon: The University of Arizona Press.
Ichiki M, Baba K, Obayashi M, and Utada H (2006) Water content and geotherm in the upper mantle above the stagnant slab: Interpretation of electrical conductivity and seismic P-wave velocity models. *Physics of the Earth and Planetary Interiors* 155: 1–15.
Ingrin J and Skogby H (2000) Hydrogen in nominally anhydrous upper-mantle minerals: Concentration levels and implications. *European Journal of Mineralogy* 12: 543–570.
Inoue T (1994) Effect of water on melting phase relations and melt composition in the system Mg_2SiO_4–$MgSiO_3$–H_2O up to 15 GPa. *Physics of Earth and Planetary Interiors* 85: 237–263.
Inoue T, Wada T, Sasaki R, and Yurimoto H (2010) Water partitioning in the Earth's mantle. *Physics of the Earth and Planetary Interiors* 183: 245–251.
Ito E, Harris DM, and Anderson AT (1983) Alteration of oceanic crust and geologic cycling of chlorine and water. *Geochemica and Cosmochemica Acta* 47: 1613–1624.
Iwamori H (1993) Dynamic disequilibrium melting model with porous flow and diffusion-controlled chemical equilibration. *Earth and Planetary Science Letters* 114: 301–313.
Jacobsen SD (2006) Effect of water on the equation of state of nominally anhydrous minerals. In: Keppler H and Smyth JR (eds.) *Water in Nominally Anhydrous Minerals*, pp. 321–342. Washington, DC: Mineralogical Society of America.
Jacobsen SD, Jiang F, Mao Z, et al. (2008) Effects of hydration on the elastic properties of olivine. *Geophysical Research Letters* 35. http://dx.doi.org/10.1029/2008GL034398.
Jacobsen SD and Smyth JR (2006) Effect of water on the sound velocities of ringwoodite in the transition zone. In: Jacobsen SD and van der Lee S (eds.) *Earth's Deep Water Cycle*, pp. 131–145. Washington, DC: American Geophysical Union.
Jacobsen SD, Smyth JR, Spetzler HA, Holl CM, and Frost DJ (2004) Sound velocities and elastic constants of iron-bearing hydrous ringwoodite. *Physics of the Earth and Planetary Interiors* 143(144): 47–56.
Jarosewich E (1990) Chemical analyses of meteorites: A compilation of stony and iron meteorite analyses. *Meteoritics* 25: 323–337.
Jarrard RD (2003) Subduction fluxes of water, carbon dioxide, chlorine, and potassium. *Geochemistry, Geophysics, Geosystems* 4. http://dx.doi.org/10.1029/2002GC000392.
Jing Z and Karato S (2011) A new approach to the equation of state of silicate melts: An application of the theory of hard sphere mixtures. *Geochimica et Cosmochimica Acta* 75: 6780–6802.
Jones AG, Lazaeta P, Fergurson IJ, et al. (2003) The electrical structure of the Slave craton. *Lithos* 71: 505–527.
Jug K, Heidelberg B, and Bredow T (2007) Cyclic cluster study on the formation of brucite from periclase and water. *Journal of Physical Chemistry* C111: 13103–13108.
Karato S (1986) Does partial melting reduce the creep strength of the upper mantle? *Nature* 319: 309–310.
Karato S (1995) Effects of water on seismic wave velocities in the upper mantle. *Proceedings of the Japan Academy* 71: 61–66.
Karato S (2006a) Influence of hydrogen-related defects on the electrical conductivity and plastic deformation of mantle minerals: A critical review. In: Jacobsen SD and van der Lee S (eds.) *Earth's Deep Water Cycle*, pp. 113–129. Washington, DC: American Geophysical Union.
Karato S (2006b) Remote sensing of hydrogen in Earth's mantle. In: Keppler H and Smyth JR (eds.) *Water in Nominally Anhydrous Minerals*, pp. 343–375. Washington, DC: Mineralogical Society of America.
Karato S (2008a) *Deformation of Earth Materials: Introduction to the Rheology of the Solid Earth*, p. 463. Cambridge: Cambridge University Press.
Karato S (2008b) Insights into the nature of plume–asthenosphere interaction from central Pacific geophysical anomalies. *Earth and Planetary Science Letters* 274: 234–240.
Karato S (2010a) The influence of anisotropic diffusion on the high-temperature creep of a polycrystalline aggregate. *Physics of the Earth and Planetary Interiors* 183: 468–472.
Karato S (2010b) Rheology of the deep upper mantle and its implications for the preservation of the continental roots: A review. *Tectonophysics* 481: 82–98.
Karato S (2011) Water distribution across the mantle transition zone and its implications for global material circulation. *Earth and Planetary Science Letters* 301: 413–423.
Karato S (2012) On the origin of the asthenosphere. *Earth and Planetary Science Letters* 321(322): 95–103.
Karato S (2013a) Geophysical constraints on the water content of the lunar mantle and its implications for the origin of the Moon. *Earth and Planetary Science Letters* 384: 144–153.
Karato S (2013b) Theory of isotope diffusion in a material with multiple-species and its implications for hydrogen-enhanced electrical conductivity in olivine. *Physics of the Earth and Planetary Interiors* 219: 49–54.

Karato S (2014a) Asymmetric shock heating and the terrestrial magma ocean origin of the Moon. *Proceedings of the Japan Academy* B90: 97–103.

Karato S (2014b) Does partial melting explain geophysical anomalies? *Physics of the Earth and Planetary Interiors* 228: 300–306.

Karato S and Jung H (2003) Effects of pressure on high-temperature dislocation creep in olivine polycrystals. *Philosophical Magazine A* 83: 401–414.

Karato S, Jung H, Katayama I, and Skemer PA (2008) Geodynamic significance of seismic anisotropy of the upper mantle: New insights from laboratory studies. *Annual Review of Earth and Planetary Sciences* 36: 59–95.

Karato S and Murthy VR (1997) Core formation and chemical equilibrium in the Earth I. physical considerations. *Physics of Earth and Planetary Interiors* 100: 61–79.

Karato S, Paterson MS, and Fitz Gerald JD (1986) Rheology of synthetic olivine aggregates: Influence of grain-size and water. *Journal of Geophysical Research* 91: 8151–8176.

Karato S and Wang D (2013) Electrical conductivity of minerals and rocks. In: Karato S (ed.) *Physics and Chemistry of the Deep Earth*, pp. 145–182. New York: Wiley-Blackwell.

Kasting JF (1989) Long-term stability of the Earth's climate. *Paleogeography Paleoclimate Paleoecology* 75: 83–96.

Kasting JF (1993) Earth's early atmosphere. *Science* 259: 920–926.

Kasting JF and Catling D (2003) Evolution of a habitable planet. *Annual Review of Astronomy and Astrophysics* 41: 429–463.

Kasting JF and Holm NG (1992) What determines the volume of ocean? *Earth and Planetary Science Letters* 109: 507–515.

Kasting JF and Pollack JB (1983) Loss of water from Venus. I. Hydrodynamic escape of hydrogen. *Icarus* 53: 479–508.

Kasting JF, Whitmire DP, and Reynolds RT (1993) Habitable zones around main sequence stars. *Icarus* 101: 108–128.

Katayama I and Karato S (2008a) Effects of water and iron content on the rheological contrast between garnet and olivine. *Physics of the Earth and Planetary Interiors* 166: 59–66.

Katayama I and Karato S (2008b) Low-temperature high-stress rheology of olivine under water-rich conditions. *Physics of the Earth and Planetary Interiors* 168: 125–133.

Katsura T, Sato K, and Ito E (1998) Electrical conductivity of silicate perovskite at lower-mantle conditions. *Nature* 395: 493–495.

Kaula WM (1979) Thermal evolution of Earth and Moon growing by planetesimal impacts. *Journal of Geophysical Research* 84: 999–1008.

Kaula WM (1990) Venus: A contrast in evolution to Earth. *Science* 247.

Kawazoe T, Ohuchi T, Nishihara Y, Nishiyama N, Fujino K, and Irifune T (2013) Seismic anisotropy in the mantle transition zone induced by shear deformation of wadsleyite. *Physics of the Earth and Planetary Interiors* 216: 91–98.

Kelbert A, Schultz A, and Egbert G (2009) Global electromagnetic induction constraints on transition-zone water content variations. *Nature* 460: 1003–1006.

Keppler H and Bolfan-Casanova N (2006) Thermodynamics of water solubility and partitioning. In: Keppler H and Smyth JR (eds.) *Water in Nominally Anhydrous Minerals*, pp. 193–230. Washington, DC: Mineralogical Society of America.

Keppler H, Wiedenbeck M, and Shcheka SS (2003) Carbon solubility in olivine and the mode of carbon storage in the Earth's mantle. *Nature* 424: 414–416.

Kleine T and Rudge JF (2011) Chronometry of meteorites and the formation of the Earth and Moon. *Elements* 7: 41–46.

Kneller EA, van Keken PE, Karato S, and Park J (2005) B-type olivine fabric in the mantle wedge: Insights from high-resolution non-Newtonian subduction zone models. *Earth and Planetary Science Letters* 237: 781–797.

Kohlstedt DL (2002) Partial melting and deformation. In: Karato S and Wenk H-R (eds.) *Plastic Deformation of Minerals and Rocks*, pp. 121–135. Washington, DC: Mineralogical Society of America.

Kohlstedt DL (2009) Properties of rocks and minerals – Constitutive equations, rheological behavior, and viscosity of rocks. In: Price GD (ed.) *Treatise on Geophysics*, pp. 389–417. Amsterdam: Elsevier.

Kohlstedt DL, Keppler H, and Rubie DC (1996) Solubility of water in the α, β and γ phases of $(Mg, Fe)_2SiO_4$. *Contributions to Mineralogy and Petrology* 123: 345–357.

Kokubo E and Ida S (2000) Formation of protoplanets from planetesimals in the solar nebula. *Icarus* 143: 15–27.

Konopliv AS, Asmar SW, Carranza E, Sjogren WL, and Yuan DN (2001) Recent gravity models as a result of the Lunar Prospector mission. *Icarus* 150: 1–18.

Korenaga J (2008) Plate tectonics, flood basalts and the evolution of Earth's ocean. *Terra Nova* 20: 419–439.

Korenaga J (2011) Thermal evolution with a hydrating mantle and the initiation of plate tectonics in the early Earth. *Journal of Geophysical Research* 116. http://dx.doi.org/10.1029/2011JB008410.

Kushiro I (2001) Partial melting experiments on peridotite and origin of mid-ocean ridge basalt. *Annual Review of Earth and Planetary Sciences* 29: 71–107.

Kushiro I, Syono Y, and Akimoto S (1968) Melting of a peridotite nodule at high pressures and high water pressures. *Journal of Geophysical Research* 73: 6023–6029.

Lainey V, Dehant V, and Pätzold M (2007) First numerical ephemerides of the Martian moons. *Astronomy and Astrophysics* 465: 1075–1084.

Lange MA and Ahrens TJ (1982) The evolution of impact-generated atmosphere. *Icarus* 51: 96–120.

Langmuir CH and Broecker WS (2012) *How to Build a Habitable Planet*, 2nd edn., p. 718. Princeton: Princeton University Press.

Larimer JW (1967) Chemical fractionations in meteorites – I. Condensation of the elements. *Geochemica et Cosmochemica Acta* 31: 1215–1238.

Lauretta DS (2011) A cosmochemical view of the solar system. *Elements* 7: 11–16.

Leahy G and Bercovici D (2007) On the dynamics of hydrous melt layer above the transition zone. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JB004631.

Lewis JS (1972) Metal/silicate fractionation in the solar system. *Earth and Planetary Science Letters* 15: 286–290.

Lewis JS (1974) Chemical composition of the solar system. *Scientific American* 230: 50–65.

Lifshitz IM and Shikin VB (1965) The theory of diffusional viscous flow of polycrystalline solids. *Soviet Physics, Solid State* 6: 2211–2218.

Lister CRB (1980) Heat flow and hydrothermal circulation. *Annual Review of Earth and Planetary Sciences* 8: 95–117.

Litasov KD and Ohtani E (2007) Effect of water on the phase relations in Earth's mantle and deep water cycle. In: Ohtani E (ed.) *Advances in High-Pressure Mineralogy*, pp. 115–156. Geological Society of America.

Lowell RP, Rona PA, and von Herzen RP (1995) Seafloor hydrothermal systems. *Journal of Geophysical Research* 100: 327–352.

Mao Z, Jacobsen SB, Jiang F, et al. (2008) Single-crystal elasticity of wadsleyite, β-Mg_2SiO_4, containing 0.37–1.66 wt.% H_2O. *Earth and Planetary Science Letters* 266: 78–89.

Marty B (2012) The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth and Planetary Science Letters* 313(314): 56–66.

Marty B and Yokochi R (2006) Water in early Earth. In: Keppler H and Smyth JR (eds.) *Water in Nominally Anhydrous Minerals*, pp. 421–450. Washington, DC: Mineralogical Society of America.

Maruyama S, Ikoma M, Genda H, Hirose K, Yokoyama T, and Santosh M (2013) The naked planet Earth: Most essential pre-requisite for the origin and evolution of life. *Geoscience Frontiers* 4: 141–165.

Matsui T and Abe Y (1986a) Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. *Nature* 319: 303–305.

Matsui T and Abe Y (1986b) Impact-induced atmospheres and oceans on Earth and Venus. *Nature* 322: 526–528.

McCarthy C, Takei Y, and Hiraga T (2011) Experimental study of attenuation and dispersion over a broad frequency range: 2. The universal scaling of polycrystalline materials. *Journal of Geophysical Research* 116. http://dx.doi.org/10.1029/2011JB008384.

McGovern PJ and Schubert G (1989) Thermal evolution of the Earth: Effects of volatile exchange between atmosphere and interior. *Earth and Planetary Science Letters* 96: 27–37.

McKenzie D (1984) The generation and compaction of partially molten rocks. *Journal of Petrology* 25: 713–765.

McMillan PF (1994) Water solubility and speciation models. In: Carroll MR and Holloway JR (eds.) *Volatiles in Magmas*, pp. 131–156. Washington, DC: Mineralogical Society of America.

McSween HYJ, Grove TL, Lentz RCF, et al. (2001) Geochemical evidence for magmatic water within Mars from pyroxenes in the Shergotty meteorite. *Nature* 409: 487–490.

Mei S and Kohlstedt DL (2000a) Influence of water on plastic deformation of olivine aggregates, 1. Diffusion creep regime. *Journal of Geophysical Research* 105: 21457–21469.

Mei S and Kohlstedt DL (2000b) Influence of water on plastic deformation of olivine aggregates, 2. Dislocation creep regime. *Journal of Geophysical Research* 105: 21471–21481.

Meier U, Trampert J, and Curtis A (2009) Global variations of temperature and water content in the mantle transition zone from higher mode surface waves. *Earth and Planetary Science Letters* 282: 91–101.

Mejias JA, Berry AJ, Refson K, and Fraser DG (1999) The kinetics and mechanism of MgO dissolution. *Chemical Physics Letters* 314: 558–563.

Miranda LF, Gómez Y, Anglada G, and Torrelles JM (2001) Water-maser emission from a planetary nebula with a magnetized torus. *Nature* 414: 284–286.

Mojzsis SJ, Harrison TM, and Pidgeon RT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409: 178–181.
Morbidelli A, Chambers JE, Lunine JI, Petit JM, Valsecchi GB, and Cyr K (2000) Source regions and timescales for the delivery of water to the Earth. *Meteoritics and Planetary Sciences* 35: 1309–1320.
Muralidharan K, Deymier P, Stimfl M, de Leeuw NH, and Drake MJ (2008) Origin of water in the inner Solar System: A kinetic Monte Carlo study of water adsorption on forsterite. *Icarus* 198: 400–407.
Mysen BO and Kushiro I (1988) Condensation, evaporation, melting, and crystallization in the primitive solar nebula: Experimental data in the system MgO–SiO_2–H_2 to 10^{-9} bar and 1870 °C with variable oxygen fugacity. *American Mineralogist* 73: 1–19.
Mysen BO, Virgo D, Popp RK, and Bertka CM (1998) The role of H_2O in Martian magmatic systems. *American Mineralogist* 83: 942–946.
Naif S, Key K, Constable S, and Evans RL (2013) Melt-rich channel observed at the lithosphere–asthenosphere boundary. *Nature* 495: 356–359.
Nakada M and Lambeck K (1991) Late Pleistocene and Holocene sea-level change: Evidence for lateral mantle viscosity variation? In: Sabadini R, Lambeck K, and Boschi E (eds.) *Glacial Isostasy, Sea Level and Mantle Rheology*, pp. 79–94. Dortrecht: Kluwer Academic.
Ni H and Keppler H (2013) Carbon in silicate melts. *Reviews in Mineralogy & Geochemistry* 75: 251–287.
Ni H, Keppler H, and Behrens H (2011) Electrical conductivity of hydrous basaltic melts: Implications for partial melting in the upper mantle. *Contributions to Mineralogy and Petrology* 162: 637–650.
Nishihara Y, Shinmei T, and Karato S (2008) Effects of chemical environments on the hydrogen-defects in wadsleyite. *American Mineralogist* 93: 831–843.
Ohtani E (1985) The primordial terrestrial magma ocean and its implications for stratifications of the mantle. *Earth and Planetary Science Letters* 78: 70–80.
Ohtani E (2005) Water in the mantle. *Elements* 1: 25–30.
Ohtani E, Hirano N, Kondo T, Ito M, and Kikegawa T (2005) Iron–water interaction at high pressure and temperature, and hydrogen transport to the core. *Physics and Chemistry of Minerals* 32: 77–82.
Ohtani E, Litasov K, Hosoya T, Kubo T, and Kondo T (2004) Water transport into the deep mantle and formation of a hydrous transition zone. *Physics of the Earth and Planetary Interiors* 143(144): 255–269.
Ohuchi T and Irifune T (2013) Development of A-type olivine fabric in water-rich upper mantle. *Earth and Planetary Science Letters* 362: 20–30.
Okuchi T (1997) Hydrogen partitioning into molten iron at high pressure: Implications for Earth's core. *Science* 278: 1781–1784.
Otsuka K and Karato S (2013) The influence of ferric iron and hydrogen on Fe–Mg interdiffusion in ferropericlase in the lower mantle. *Physics and Chemistry of Minerals,* in press.
Otsuka K, Longo M, McCammon CA, and Karato S (2013) Ferric iron content of ferropericlase as a function of composition, oxygen fugacity, temperature and pressure: Implications for redox conditions during diamond formation in the lower mantle. *Earth and Planetary Science Letters* 365: 7–16.
Ozima M and Igarashi G (1989) Terrestrial noble gases: Constraints and implications on atmospheric evolution. In: Atreya SK, Pollack JB, and Matthews MS (eds.) *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 306–327. Tuscon: The University of Arizona Press.
Pahlevan K and Stevenson DJ (2007) Equilibration in the aftermath of the lunar-forming giant impact. *Earth and Planetary Science Letters* 262: 438–449.
Parai R and Mukhopadhyay S (2012) How large is the subducted water flux? New constraints on mantle regassing rates. *Earth and Planetary Science Letters* 317–318: 396–406.
Pearson DG, et al. (2014) Hydrous mantle transition zone indicated by ringwoodite included within diamond. *Nature* 507: 221–224.
Pepin RO (1997) Evolution of Earth's noble gases: Consequence of assuming hydrodynamic loss driven by giant impact. *Icarus* 126: 148–156.
Peslier AH (2010) A review of water contents of nominally anhydrous natural minerals in the mantles of Earth, Mars and the Moon. *Journal of Volcanology and Geothermal Research* 197: 239–258.
Peslier AH, Woodland AB, Bell DR, and Lazarov M (2010) Olivine water contents in the continental lithosphere and the longevity of cratons. *Nature* 467: 78–83.
Peyronneau J and Poirier J-P (1989) Electrical conductivity of the material of the Earth's lower mantle. *Nature* 365: 537–539.
Post AD, Tullis J, and Yund RA (1996) Effects of chemical environment on dislocation creep of quartzite. *Journal of Geophysical Research* 101: 22143–22155.
Pritchard ME and Stevenson DJ (2000) Thermal aspects of a lunar origin by giant impact. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tuscon: The University of Arizona Press.
Raj R (1975) Transient behavior of diffusion-induced creep and creep rupture. *Metallurgical Transactions* 6A: 1499–1509.
Raj R and Ashby MF (1971) On grain boundary sliding and diffusional creep. *Metallurgical Transactions* 2: 1113–1127.
Ranero CR, Phipps Morgan J, McIntosh KD, and Reichert C (2003) Bending, faulting, and mantle serpentinization at the Middle America trench. *Nature* 425: 367–373.
Ray RD, Eanes RJ, and Lemoine FG (2001) Constraints on energy dissipation in the Earth's body tide from satellite tracking and altimetry. *Geophysical Journal International* 144: 471–480.
Raymond SN, Quinn T, and Lunine JI (2007) High-resolution simulations of the final assembly of Earth-like planets. 2. Water delivery and planetary habitability. *Astrobiology* 7: 66–84.
Reymer A and Schubert G (1984) Phanerozoic addition rates to the continental crust and crustal growth. *Tectonics* 3: 63–77.
Ribe NM (1985) The generation and compaction of partial melts in the earth's mantle. *Earth and Planetary Science Letters* 73: 361–376.
Richard G, Monnereau M, and Ingrin J (2002) Is the transition zone an empty water reservoir? Inference from numerical model of mantle dynamics. *Earth and Planetary Science Letters* 205: 37–51.
Richards MA, Bunge H-P, and Lithgow-Bertelloni C (2000) Mantle convection and plate motion history: Toward general circulation models. In: Richards MA, Gordon RG, and Hilst RDvd (eds.) *The History and Dynamics of Global Plate Motions*, pp. 289–308. Washington, DC: American Geophysical Union.
Richter FM and McKenzie D (1984) Dynamical models for melt segregation from a deformable matrix. *The Journal of Geology* 92: 729–740.
Rikitake T (1966) *Electromagnetism and the Earth's Interior*. Amsterdam: Elsevier, 308 pp.
Ringwood AE (1975) *Composition and Structure of the Earth's Mantle,* p. 618. New York: McGraw-Hill.
Ringwood AE (1979) *Origin of the Earth and Moon*, p. 295. Berlin: Springer-Verlag.
Ritsema J, Deuss A, van Heijst HJ, and Woodhouse JH (2011) S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International* 184: 1223–1236.
Robert F, Gautier D, and Dubrulle B (2003) The solar system D/H ratio: Observations and theories. *Space Science Reviews* 92: 201–224.
Rohrbach A and Schmidt MW (2011) Redox freezing and melting in the Earth's mantle resulting from carbon–iron redox coupling. *Nature* 472: 209–214.
Romanowicz B and Durek JJ (2000) Seismological constraints on attenuation in the Earth: A review. In: Karato S, Forte AM, Liebermann RC, Masters G, and Stixrude L (eds.) *Earth's Deep Interior*, pp. 161–179. Washington, DC: American Geophysical Union.
Ross JV and Nielsen KC (1978) High-temperature flow of wet polycrystalline enstatite. *Tectonophysics* 44: 233–261.
Rubey W (1951) Geologic history of sea water, an attempt to state the problem. *Geological Society of America Bulletin* 62: 1111–1148.
Rubie DC, Nimmo F, and Melosh HJ (2007) Formation of Earth's core. In: Stevenson D (ed.) *Evolution of the Earth*, pp. 51–90. Amsterdam: Elsevier.
Rüpke LH, Phipps Morgan J, and Dixon JE (2006) Implications of subduction rehydration for Earth's deep water cycle. In: Jacobsen SD and Lee Svd (eds.) *Earth's Deep Water Cycle*, pp. 263–276. Washington, DC: American Geophysical Union.
Rüpke LH, Phipps Morgan J, Hort M, and Connolly JAD (2004) Serpentine and the subduction zone water cycle. *Earth and Planetary Science Letters* 223: 17–34.
Saal AE, Hauri EH, Van Orman JA, and Rutherford MC (2013) Hydrogen isotopes in lunar volcanic glasses and melt inclusions reveal a carbonaceous chondrite heritage. *Science* 340: 1317–1320.
Safronov VS (1972) *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets,* p. 212. Jerusalem: Keter Publishing House.
Safronov VS (1978) The heating of the earth during its formation. *Icarus* 33: 3–12.
Sarda P, Staudacher T, and Allègre CJ (1985) $^{40}Ar/^{36}Ar$ in MORB glasses: Constraints on atmosphere and mantle evolution. *Earth and Planetary Science Letters* 72: 357–375.
Sasselov DD and Lecar M (2000) On the snow line in dusty protoplanetary disks. *The Astrophysical Journal* 528: 995–998.
Savage MK (1999) Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Review of Geophysics* 37: 65–106.
Schmandt B, Dueker K, Hansen SM, Jabinsek JJ, and Zhang Z (2011) A sporadic low-velocity layer atop the western U.S. mantle transition zone and short-wavelength variations in transition zone discontinuities. *Geochemistry, Geophysics, Geosystems* 12. http://dx.doi.org/10.1029/2011GC003668.
Schmidt MW and Poli S (1998) Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth and Planetary Science Letters* 163: 361–379.

Shankland TJ, Peyronneau J, and Poirier J-P (1993) Electrical conductivity of the Earth's lower mantle. *Nature* 399: 453–455.

Shcheka SS, Wiedenbeck M, Frost DJ, and Keppler H (2006) Carbon solubility in mantle minerals. *Earth and Planetary Science Letters* 245: 730–742.

Shearer PM (2000) Upper mantle discontinuities. In: Karato S, Forte AM, Liebermann RC, Masters G, and Stixrude L (eds.) *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scales*, pp. 115–131. Washington, DC: American Geophysical Union.

Shibazaki Y, et al. (2012) Sound velocity measurements in dhcp-FeH to 70 GPa with inelastic X-ray scattering: Implications for the composition of the Earth's core. *Earth and Planetary Science Letters* 313(314): 79–85.

Shito A, Karato S, Matsukage KN, and Nishihara Y (2006) Toward mapping water content, temperature and major element chemistry in Earth's upper mantle from seismic tomography. In: Jacobsen SD and Lee Svd (eds.) *Earth's Deep Water Cycle*, pp. 225–236. Washington, DC: American Geophysical Union.

Shu FH, Adams FC, and Lizano S (1987) Star formation in molecular clouds: Observations and theory. *Annual Review of Astronomy and Astrophysics* 25: 23–81.

Sifré D, Gardés E, Massuyeau M, Hashin L, Hier-Majumder CA, and Gaillard F (2014) Electric conductivity during incipient melting in the oceanic low-velocity zone. *Nature* 509: 81–85.

Silver PG (1996) Seismic anisotropy and mantle deformation: Probing the depths of geology. *Annual Review of Earth and Planetary Sciences* 24: 385–432.

Sleep NH (1988) Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophysical Journal of Royal Astronomical Society* 95: 437–447.

Sleep NH (1992) Archean plate tectonics: What can be learn from continental geology? *Canadian Journal of Earth Sciences* 29: 2066–2071.

Sleep NH (2001) Initiation of clement surface conditions on the earliest Earth. *Proceedings of National Academy of Sciences* 98: 3666–3672.

Smyth JR and Frost DJ (2002) The effect of water on the 410-km discontinuity: An experimental study. *Geophysical Research Letters* 29. http://dx.doi.org/10.129/2001GL014418.

Solomatov VS (2009) Magma oceans and primordial mantle differentiation. In: Stevenson DJ (ed.) *Evolution of the Earth*, pp. 91–119. Amsterdam: Elsevier.

Som SM, Catling DC, Harnmeijer JP, Polivka PM, and Buick R (2012) Air density 2.7 billion years ago limited to less than twice modern levels by fossil raindrop imprints. *Nature* 484: 359–362.

Stein M and Ben-Avraham Z (2007) Mechanisms of continental crust growth. In: Stevenson D (ed.) *Evolution of the Earth*, pp. 171–195. Amsterdam: Elsevier.

Stevenson DJ (1987) Origin of the moon—The collision hypothesis. *Annual Review of Earth and Planetary Sciences* 15: 271–315.

Stevenson DJ (1990) Fluid dynamics of core formation. In: Jones HENaJH (ed.) *Origin of the Earth*, pp. 231–249. Oxford: Oxford University Press.

Suetsugu D, Inoue T, Yamada A, Zhao D, and Obayashi M (2006) Towards mapping three-dimensional distribution of water in the transition zone from P-wave velocity tomography and 660-km discontinuity depths. In: Jacobsen SD and Lee Svd (eds.) *Earth's Deep Water Cycle*, pp. 237–249. Washington, DC: American Geophysical Union.

Takei Y (2002) Effect of pore geometry on Vp/Vs: From equilibrium geometry to crack. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000522.

Takei Y and Holtzman B (2009) Viscous constitutive relations of solid–liquid composites in terms of grain boundary contiguity: 1. Grain boundary diffusion control model. *Journal of Geophysical Research* 114. http://dx.doi.org/10.1029/2008JB005850.

Tarits P, Hautot S, and Perrier F (2004) Water in the mantle: Results from electrical conductivity beneath the French Alps. *Geophysical Research Letters* 31. http://dx.doi.org/10.1029/2003GL019277.

Tauzin B, Debayle E, and Wittingger G (2010) Seismic evidence for a global low-velocity layer within the Earth's upper mantle. *Nature Geoscience* 3: 718–721.

Tonks WB and Melosh HJ (1990) The physics of crystal settling and suspension in a turbulent magma ocean. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 151–174. New York: Oxford University Press.

Turcotte DL and Schubert G (1982) *Geodynamics: Applications of Continuum Physics to Geological Problems*, p. 450. New York: John Wiley & Sons.

Tyburczy JA, Frisch B, and Ahrens TJ (1986) Shock-induced volatile loss from a carbonaceous chondrite – Implications for planetary accretion. *Earth and Planetary Science Letters* 80: 201–207.

Tyburczy JA, Xu XM, Ahrens TJ, and Epstein S (2001) Shock-induced devolatilization and isotopic fractionation of H and C from Murchison meteorite: Some implications for planetary accretion. *Earth and Planetary Science Letters* 192: 23–30.

Ulmer P and Trommsdorff V (1995) Serpentine stability to mantle depths and subduction-related magmatism. *Science* 268: 858–861.

Urey HC (1952) *The Planets: Their Origin and Development*, p. 245. New Haven: Yale University Press.

Van Avendonk HJA, Holbrook WS, Lizarralde D, and Denyer P (2011) Structure and serpentinization of the subducting Cocos plate offshore Nicaragua and Costa Rica. *Geochemistry, Geophysics, Geosystems* 12. http://dx.doi.org/10.1029/2011GC003592.

van Keken PE, Hacker BR, Syracuse EM, and Abers GA (2011) Subduction factory 4: Depth-dependent flux of H_2O from subducting slabs worldwide. *Journal of Geophysical Research* 116. http://dx.doi.org/10.1029/2010JB007922.

van Thienen P, Benzerara K, Breuer D, et al. (2007) Water, life, and planetary geodynamical evolution. *Space Science Reviews* 129: 167–203.

Vattuone L, Smerieri M, Savio L, et al. (2013) Accretion disk origin of the Earth's water. *Philosophical Transactions of the Royal Society of London* A371. http://dx.doi.org/10.1098/rsta.2011.0585.

Walker JCG, Hays PB, and Kasting JF (1981) A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86: 9776–9782.

Wallmann K (2001) The geological water cycle and the evolution of marine $\delta^{18}O$ values. *Geochimica et Cosmochimica Acta* 65: 2469–2485.

Walter MJ, Kohn SC, Araujo D, et al. (2011) Deep mantle cycling of oceanic crust: Evidence from diamonds and their mineral inclusions. *Science* 334: 54–57.

Wang D, Karato S, and Jiang Z (2013) An experimental study of the influence of graphite on the electrical conductivity of olivine aggregate. *Geophysical Research Letters* 40: 2028–2032.

Wang D, Mookherjee M, Xu Y, and Karato S (2006) The effect of water on the electrical conductivity in olivine. *Nature* 443: 977–980.

Wang W and Takahashi E (2000) Subsolidus and melting experiments of K-doped peridotite KLB-1 to 27 GPa; Its geophysical and geochemical implications. *Journal of Geophysical Research* 105: 2855–2868.

Watson EB and Harrison TM (2005) Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science* 308: 841–844.

Wetherill GW (1990) Formation of the Earth. *Annual Review of Earth and Planetary Sciences* 18: 205–256.

Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.

Williams Q and Hemley RJ (2001) Hydrogen in the deep Earth. *Annual Review of Earth and Planetary Sciences* 29: 365–418.

Wise DU (1974) Continental margins, freeboard and the volumes of continents and ocean through time. In: Burk CA and Drake CL (eds.) *Geology of Continental Margins*, pp. 45–58. New York: Springer.

Wood BJ and Halliday AN (2010) The lead isotopic age of the Earth can be explained by core formation alone. *Nature* 465: 767–771.

Wood BJ, Halliday A, and Rehkämper M (2010) Volatile accretion history of the Earth. *Nature* 567: E6–E7.

Wood BJ, Pawley AR, and Frost DR (1996) Water and carbon in the Earth's mantle. *Philosophical Transaction of Royal Society of London* 354: 1495–1511.

Worsley TR, Nance D, and Moody JB (1984) Global tectonics and eustasy for the past 2 billion years. *Marine Geology* 58: 373–400.

Wyllie PJ and Huang WL (1976) Carbonation and melting reactions in the system CaO–MgO–SiO_2–CO_2 at mantle pressure with geophysical and petrological applications. *Contributions to Mineralogy and Petrology* 54: 79–107.

Yang X (2012) Orientation-related electrical conductivity of hydrous olivine, clinopyroxene and plagioclase and implications for the structure of the lower continental crust and uppermost mantle. *Earth and Planetary Science Letters* 317(318): 241–250.

Yang X, Keppler H, McCammon C, and Ni H (2012) Electrical conductivity of orthopyroxene and plagioclase in the lower crust. *Contributions to Mineralogy and Petrology* 163: 33–48.

Yoneda S and Grossman L (1995) Condensation of CaO–MgO–Al_2O_3–SiO_2 liquids from cosmic gases. *Geochimica et Cosmochimica Acta* 59: 3413–3444.

Yoshino T, Nishihara Y, and Karato S (2007) Complete wetting of olivine grain-boundaries by a hydrous melt near the mantle transition zone. *Earth and Planetary Science Letters* 256: 466–472.

Yoshino T, Yamazaki D, and Mibe K (2009) Well-wetted olivine grain boundaries in partially molten peridotite in the asthenosphere. *Earth and Planetary Science Letters* 283: 167–173.

Young TE, Green HW II, Hofmeister AM, and Walker D (1993) Infrared spectroscopic investigation of hydroxyl in β-$(Mg, Fe)_2SiO_4$ and coexisting olivine: Implications for mantle evolution and dynamics. *Physics and Chemistry of Minerals* 19: 409–422.

9.06 Evolution of the Earth: Plate Tectonics Through Time

NH Sleep, Stanford University, Stanford, CA, USA

© 2015 Elsevier B.V. All rights reserved.

9.06.1	**Introduction**	145
9.06.2	**Physical Preliminaries**	147
9.06.2.1	Global Heat Balance	147
9.06.2.2	Heat Transfer by Seafloor Spreading	149
9.06.2.3	Parameterized Convection	149
9.06.2.4	Changes in the Mode of Convection	151
9.06.2.5	Pressure-Release Melting at Mid-Ocean Ridges	152
9.06.3	**Aftermath of the Moon-Forming Impact**	154
9.06.3.1	Early Mush Ocean	155
9.06.3.2	Steady-State Mush Ocean and Transition to Plate Tectonics	156
9.06.3.3	Zircon Evidence	157
9.06.3.4	Asteroid Bombardment	159
9.06.4	**Dawn of Plate Tectonics**	159
9.06.4.1	Seafloor Spreading and Oceanic Crust	159
9.06.4.2	Transform Faults	160
9.06.4.3	Subduction and Continental Lithosphere	160
9.06.4.4	Subduction and Continental Crust	161
9.06.5	**The Rate of Plate Tectonics over Time**	162
9.06.5.1	Longevity of Geologic Features	162
9.06.5.2	Degassing and Heat Balance	162
9.06.5.3	Freeboard and Stress Within Plates	163
9.06.6	**Death of Plate Tectonics**	163
9.06.7	**Biological Implications**	163
9.06.7.1	Origin of Life and Its Survival on the Early Earth	164
9.06.7.2	Biological Effects on Geologic Processes	164
9.06.7.3	The Extraterrestrial Export of Biological Concepts and the Future of Life	165
9.06.8	**Conclusions and Musings**	165
Appendix A	**Thermal Models**	167
Appendix B	**Convection Beneath Free-Slip Lid**	167
Acknowledgments		168
References		168

9.06.1 Introduction

The basic issues of the nature of tectonics on the early Earth have not been resolved over the last 6 years since the previous volume of the Treatise. Geologists and geochemists have continued their analyses of ancient rocks and detrital zircons. Differences between Hadean and Archean rocks with respect to later rocks have become more evident (Arndt and Nisbet, 2012; Herzberg and Rudnick, 2012; Hoffmann et al., 2011; Kemp et al., 2010; Næraa et al., 2012). Improvement of computers has led to more realistic geodynamic modeling. In particular, the causes of one-sided subduction are better understood (Crameri et al., 2012), allowing global geochemical history models to presume its existence. O'Neil et al. (2008, 2011, 2012) and Adam et al. (2012) studied Hadean greenstone and intrusive gabbro in the Nuvvuagittuq region of Quebec. Detection of antineutrinos from radioactive decay of Th and U has the potential to constrain heat production in the Earth's interior (Dye, 2012; Mareschal et al., 2012; Šrámek et al., 2012).

Nonincremental progress continues from biology, as molecular studies of the origin of life have become explicit in terms of geologic and tectonic environment (Lane et al., 2010; Mulkidjanian et al., 2012; Russell et al., 2010). The task from Earth science is a mixture of vetting the reasonability of these putative environments, using firmer biological constraints on ancient tectonic and environmental processes and providing biologists a catalog of potentially habitable environments (e.g., Arndt and Nisbet, 2012; Boussau and Gouy, 2012). In particular, the Earth teemed with life by 3.85 billion years ago (Ga) implying that the surface climate was cool enough for life. Thus, a massive >25 bar CO_2 greenhouse no longer existed at that time, and most of the Earth's CO_2 had already subducted into the mantle. By then, life had already strongly affected the chemistry of the crust and even the mantle (Foley, 2011; Hazen et al., 2008; Papineau, 2010; Sleep, 2010; Sleep et al., 2012). The possibility that life has modulated global tectonic processes thus arises.

The scientific issues that arose during the advent of the theory of plate tectonics still serve to place preservation problems on the modern issue of tectonics on the early Earth into context. The existence of seafloor spreading in the Atlantic and ridge–ridge transform faults was established when the author

attended his first international scientific conference at Woods Hole in the summer of 1967. Later, the approximation of rigid plates on a spherical shell provided the global geometry, taking advantage of transform faults. Subduction provided an explanation of deep-focus earthquakes and the sink for the oceanic lithosphere produced by seafloor spreading. It was evident at the Woods Hole meeting that no one there had any idea how seafloor spreading works. This lacuna did not greatly concern the mainly observational scientists in attendance.

The initial hypothesis explained continental drift, relegating continents to plate passengers. Atwater (1970) brought the concept ashore in California demonstrating its great explicative power to continental tectonics. Overall, the plate tectonic revolution shifted attention away from cratons. Yet the interior of continents had geologic records of real events.

In December 1975, the author attended the Penrose Conference on continental interiors, the final stationary bastion under siege. Well-organized fixists hijacked the conference, jettisoned the published program, and mounted a determined counterattack. Speaker after speaker droned on that his locality was very complicated, that plate tectonics was of no use in interpreting the geology, that it certainly could not be established just looking at his quadrangle, and thus that the theory is dead wrong.

With regard to tectonics in the Earth's deep past, we are much like land-based geologists in 1975. The biases of geologic preservation place us in the same bind that provincial data selection placed the Penrose holdouts during that freezing December in San Diego. We have little intact seafloor older than the ~180 Ma crust along the passive margins of the central Atlantic (Moores, 2002). We cannot observe the linchpins of modern plate tectonics, magnetic stripes, and ridge–ridge transform faults. We certainly cannot observe earthquake mechanisms, the geoid, and tomography. Essentially, we are stuck with data from continents, the very regions that befuddled earlier geologists.

Yet we can do better with the deep past than Earth science did with the Tertiary in 1960. We can view geology at appropriate scales with the caveat that later tectonics dispersed the Archean record into ~35 blocks (Bleeker, 2003). We have a reasonable understanding of the kinematics of modern Earth, especially the record-producing process of continental breakup eventually followed by continental collision. Geodynamicists picked much of the low-hanging fruit from the orchard opened by plate tectonics over the last 50 years. We know how the modern Earth works well enough that we have hope of exporting concepts in time and to other planets.

Export inflicts some discussion of semantics. Plate tectonics, as originally conceived, was an approximate kinematic theory. It unified continental drift, seafloor spreading, subduction, and transform faulting. Orogeny from continental collision provided an apt explanation of mountain belts. These processes are linked but somewhat disjoint. For example, seafloor spreading does not kinematically require narrow ridge axes, that the plates are rigid, that subduction is one-sided, or even that transform faults exist. Mantle plumes and thick cratonal lithosphere are significant features on the modern Earth that are further afield from plate processes.

With regard to the cardinal kinematic postulate, rigid plates tessellate 85% of the Earth's surface (e.g., Zatman et al., 2005). Diffuse plate boundaries grout the remainder. Diffuse deformation zones divide major oceanic plates into subplates. Typically, the pole of rotation for the two subplates lies near their boundary so that both extension and compression occur. This process may nucleate subduction and does cause great earthquakes (McGuire and Beroza, 2012) but is essentially lost in the fog of time. Diffuse continental zones, like the Basin and Range, the Lena River region of Siberia, and Tibet, occur within hot weak continental crust. Both extension and compression have high preservation potential including sediments, focusing attention on vertical tectonics.

Continental collisions, like India with Tibet, produce strike-slip faults that extend far inland from the point of indention (Molnar and Tapponnier, 1975). Kuiper et al. (2011) described analogous escape tectonics at ~1.83 Ga on the northwest margin of the Archean Superior craton. Such faults disrupt the orogen as well as otherwise stable regions. The record has a high preservation potential that is readily associated with the plate processes.

In all cases, vertical tectonics is more easily recognized than strike-slip tectonics. This is especially true with continent-margin-parallel faults, like the San Andreas Fault and the fault system through the Sumatra volcanic arc. Lateral offset is not obvious because the fault runs along the gross geologic strike. Piercing points may be quite distant. Recently uplifted mountain ranges (e.g., in western California) are more evident than the strike-slip component of the tectonics. Historically, this view seemed natural; the US Geological Survey topographic map of Palo Alto perpetuates the archaic term the San Andreas Rift.

The question of the antiquity of plate processes arose soon after geologists accepted the modern process (Hurley and Rand, 1969). A significant literature exists. The reader is referred to the works of Sleep (1979, 1992) and Sleep and Windley (1982) for discussions of early work. Stevenson (2003) and Korenaga (2006) discussed ancient global dynamics. Bleeker (2003), Sleep (2005), Condie and Benn (2006), Polat and Kerrich (2006), and Van Kranendonk (2010) reviewed the rock record on continents. van Hunen and Moyen (2012) reviewed the dynamics of and the evidence for Archean subduction. Moores (2002) reviewed possible exposures of ancient oceanic crust. On the other hand, Stern (2005) argued that modern plate processes began in Neoproterozoic time.

Most, though not all, recent discussions of the geology and dynamics of the early Earth at least utilize plate processes as a framework to start discussion. The author winnows a subset of the work that bears on the problem at hand, begins with convection to put the physics in context, and then moves forward in time from the Earth's formation for the first evidence of various aspects of plate tectonics. To some extent, this effort involves finding least astonishment hypotheses; at worst, grabbing at straws. He then considers changes in the more recent past, finally considering the fate of the Earth in the geologic future.

Note that the geologic timescale is not yet standardized for the early Earth. Goldblatt et al. (2009) discussed an event-based provisional timescale. The Chaotan Era applies to the entire solar system and extends from the start of the solar system to the Moon-forming impact. Their subsequent timescale that applies only to the Earth is the subject of this chapter. The Hadean era begins with the Moon-forming impact.

A highly unproductive moving target results if one defines the start of the Archean as the age of the oldest rock; then, a Hadean rock can never be found. Our definition of ~3.9 Ga retains an Archean age for the oldest known sedimentary rocks in Greenland and nearby Canada.

This chapter relates to many topics in this *Treatise on Geophysics*. Volumes 6 and 7 relate to the current state of the Earth. In this volume, **Chapters 9.01**, **9.02**, **9.04**, **9.07** and **9.10** cover topics that are intimately coupled with this chapter.

9.06.2 Physical Preliminaries

Plate tectonics is a form of convection where hot material rises at ridges and cold dense slabs sink. Much of geodynamics involves this heat and mass transfer process. This chapter reviews the physics in this section so that it can be referred back when the Earth's history is discussed.

Since the nature of convection within the Earth changed over time, we loosely divide early the Earth's history into the period immediately after the Earth's formation, part of the Hadean where a 'magma' ocean filled with mostly crystalline 'mush' covered the planet, and the Archean from ~3.9 to 2.5 Ga, when it becomes increasingly meaningful to compare theory and observation. After the Archean, we can apply analogies to modern plate processes with more confidence to a reasonable record.

The Earth's mantle is complicated. Plate boundaries fail by faulting. The interior creeps as a very viscous fluid. Partial melting produces buoyant oceanic crust. These processes are complex enough that examination of the rock record is essential. It is useful to see the bases of common physical arguments about the tectonics of the Earth. This chapter begins with a discussion on the heat balance in the mantle.

9.06.2.1 Global Heat Balance

One can construct a thermal history of the Earth's interior by equating surface heat flow with contributions from transient cooling and radioactive heat generation. Formally, the heat balance for the mantle is

$$4\pi r_E^2 q = \frac{4\pi\left(r_E^3 - r_C^3\right)}{3}\rho H - \frac{4\pi r_E^3}{3}\rho C\frac{\partial T}{\partial t} + 4\pi r_C^2 q_C \qquad [1]$$

where q is the heat flow from the mantle, r_E is the Earth's radius, q_C is the heat flow at the core's radius r_C, ρ is density, C is specific heat (essentially at constant pressure), H is radioactive heat generation per mass, T is the interior mantle temperature, and t is time. We make an approximation that the core cools at the same rate as the mantle in example calculations. About 1/5 of the Earth's heat capacity is in the core, so its details can be ignored in some qualitative arguments but certainly not those related to plumes.

The specific heat of the mantle is about $1.25\ \mathrm{J\ K^{-1}\ kg^{-1}}$. This implies that a global heat flow of $1\ \mathrm{W\ m^{-2}}$ cools the mantle at $3.2\ \mathrm{K\ my^{-1}}$. Heat flows greater than this quickly cool the mantle and cannot have been sustained through even the Hadean. The present mantle heat flow of $0.070\ \mathrm{W\ m^{-2}}$ would cool the mantle at $225\ \mathrm{K\ BY^{-1}}$ and the whole Earth at $180\ \mathrm{K\ BY^{-1}}$. A major task in geodynamics is to obtain the average surface heat flow as a function of the interior temperature, which leaves radioactivity as the unknown quantity in eqn [1].

This chapter presents the results of simple thermal models of the Earth's history at this stage for use throughout this chapter. The intent is to put potentially geologically observable features and geodynamics in a common context. The models conserve energy using eqn [1].

The models are formulated so that the laterally averaged heat flow is nonmonotonic. There have been several geodynamic (Davies, 1995; Korenaga, 2008a,b; Lenardic and Crowley, 2012; O'Neill et al., 2007; Sleep, 2000; Stevenson, 2003) and petrologic (Herzberg and Rudnick, 2012; Herzberg et al., 2007, 2010) discussions of this possibility. The possibility that the globally averaged heat flow is a multibranched function of the interior temperature, which is bistable, may have been 'pondered' but has not been explicitly suggested outside of the previous version of this chapter and the work of Sleep (2000) (Lenardic and Crowley, 2012). **Figure 1** illustrates the basic properties of a bistable system. The computed

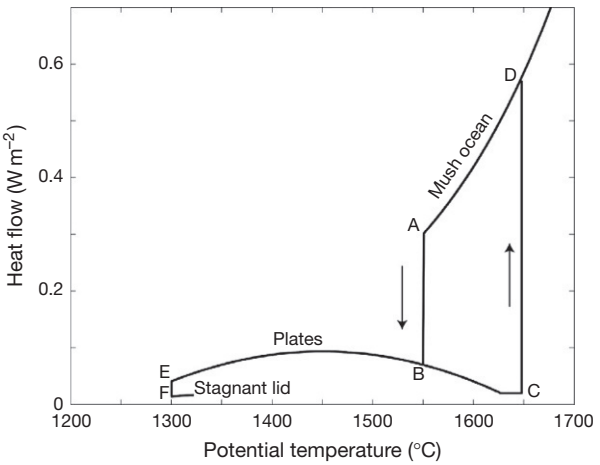


Figure 1 The heat flow from the Earth's mantle is a multibranched function of the potential temperature of the Earth's interior. Thermal histories are paths on this graph. One model has a monotonic thermal history where the heat flow lies along the transition in branch jumps. The other model jumps between branches and has a nonmonotonic thermal history. Both models start at 2000 °C at 4.5 Ga. The mantle cools along the mush ocean path until it reaches point A. The nonmonotonic model jumps to plate tectonics at point B. If the heat flow is greater than the radioactive heat generation, the mantle cools along the plate branch moving toward point E. If the heat flow is lower than the radioactive heat generation, the mantle heats up to point C, where convection jumps to the mush ocean branch at point D. It then cools to point A. The nonmonotonic model jumps to stagnant-lid convection from point E to F. The mantle then heats up. The monotonic model stays at the jumps with the heat flow equal to radioactive heat generation until radioactive heat generation decreases to heat flow of the cooler mode. Note that Thom (1983) popularized this presentation method in his catastrophe theory. I eschew the word 'catastrophe' in the text as it is preempted with a foul stench in the Earth sciences. One could add more another axes perpendicular to the plane of the diagram. Continental area (Lenardic et al., 2005) and hydration of the mantle Korenaga (2011) would be good choices. One would then need the net growth (shrinkage) age of continents and hydration as a function of interior temperature, convection mode, continental area, and hydration state to compute forward models.

thermal evolution of the Earth is a trajectory of this graph. The models generate physically realizable thermal histories, not cartoons. See **Appendix A** for the parameters used to compute these semischematic graphs.

Radioactive decay and the cooling of the Earth's interior over time are now comparable items in the Earth's heat budget. Example calculations illustrate the basic point that radioactive heat production by itself does not require very high surface heat flow on the early Earth. We thus select a value of the current heat flow supplied by radioactivity at the upper end of recent estimates that is half the total mantle heat flow (i.e., 0.035 W m^{-2}). (We use extra digits throughout this chapter to make calculations easier to follow.) For reference, the bulk silicate Earth model has 0.024 W m^{-2} in the current mantle and 0.038 W m^{-2} in current crust plus mantle (McDonough and Sun, 1995). Lyubetskaya and Korenaga (2007) and Korenaga (2008a,b) gave 19 W m^{-2} as their preferred mantle value. Šrámek et al. (2012) reviewed estimates in this range of crustal and mantle radioactivity. Antineutrino detectors soon will provide hard data on the absolute abundance of U and Th in the Earth's interior (Araki et al., 2005; Dye, 2012; Enomoto et al., 2007; Fiorentini et al., 2005; Mareschal et al., 2012; Šrámek et al., 2012). Previous cosmochemical and geochemical approaches to obtain these abundances then can be reappraised, shedding light on the formation and history of the Earth.

The heat flow from radioactivity at 4 Ga was a factor of a few more than present (i.e., ~0.13 W m^{-2} in the mantle examples presented here) (**Figure 2**). (Negative ages are in the future.) About one-third of the radioactivity is currently at shallow depths within continental crust where it does not drive mantle convection (McDonough and Sun, 1995). Continent-like chemical reservoirs existed as early as ~4.5 Ga (Bell et al., 2011; Harrison et al., 2005), but the ancient fraction of radioactivity sequestered within continents in the deep past is unknown. We use the current partition in examples, as the exact amount does not affect the gist of the arguments (**Figure 2**). For simplicity, we also ignore the direct effect of continental area on the vigor of convection (Lenardic et al., 2005).

Studies of mantle-derived igneous rocks constrain the thermal history of the Earth's interior since the early Archean. There is general agreement that the mantle was hotter at 3 Ga than now. For example, Abbott et al. (1994) and Galer and Mezger (1998) gave their preferred temperature decrease of 150 K for normal mantle. Herzberg et al. (2010) calibrated their model in **Figure 3** using a suite of screened samples. The monotonic and nonmonotonic models are also acceptable in light of their data.

The task of inferring past mantle temperature for lavas is not straightforward. The major difficulty is that the modern source temperature of mantle-derived lavas is variable, with plume-related material being hotter. Plume-derived magmas erupted at least from the Archean to the present. There has also been progress in distinguishing arc from mid-oceanic lavas in the Archean (Herzberg and Rudnick, 2012; Herzberg et al., 2010; Polat and Kerrich, 2006). A second potential difficulty is that hydrous magma melts at a lower temperature than anhydrous magma. Grove and Parman (2004) contended that komatiites are hydrous magmas and the mantle was only ~100 K hotter. Arndt (2003) noted that this is unlikely with komatiites in general.

Archean cratonal xenoliths provide additional evidence of the temperature in the Archean mantle. For example, Canil (2004) showed that trace elements in his samples are compatible with formation within a subduction zone, but not with formation from a hotter magma at their present lithospheric depth.

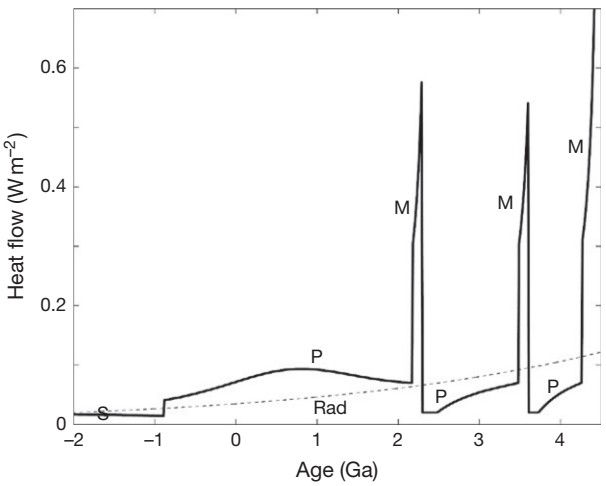


Figure 2 Computed heat flow as a function of age for the nonmonotonic model (dark line) and the radioactive heat generation as heat flow (thin dashed line). The modes of convection are (M) mush ocean, (P) plates, and (S) stagnant lid. The model resolves the expected duration of episodes, but no attempt was made to fine-tune the timing of transitions.

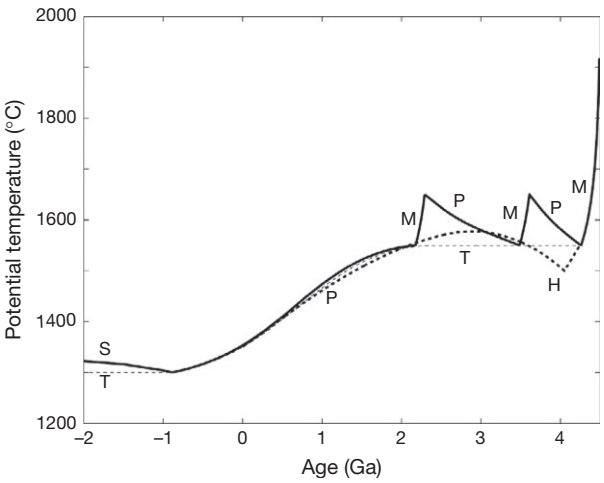


Figure 3 The computed thermal history for the nonmonotonic model (dark line) and the monotonic model (thin dashed line). The modes of convection are (M) mush ocean, (P) plates, (S) stagnant lid, and (T) transition in the monotonic model. The temperature remains constant in the transitions in the monotonic model. Herzberg et al. (2010) (H. thick dashed line) calculated the thermal history with models calibrated by inferred source region temperatures of mantle-derived igneous rocks. The curve assumes that the Urey ratio between current mantle radioactive heat generation and mantle heat flow is 0.38. (One may rescale the curves by adding or subtracting a small constant temperature.) Herzberg et al. (2010) did not compute the thermal evolution immediately after the Moon-forming impact. We extrapolate our mush ocean cooling rate to join their curve to illustrate the gross effect of overcooling.

9.06.2.2 Heat Transfer by Seafloor Spreading

The global heat balance provides necessary attributes for whatever type of global tectonics existed in the later Hadean and the Archean. Radioactive heat production was small enough that plate tectonics is a viable candidate mechanism. As at present, one needs to consider the main mechanism of heat transfer, the formation, and cooling of oceanic lithosphere.

One obtains simple well-known formulas by considering the oceanic lithosphere to be a half-space (e.g., Turcotte and Schubert, 2002). We present their first-order forms in one place so that we may refer back to them later. The roof of a liquid or mush-filled magma ocean might well cool as a plate above an isothermal half-space.

We use global spreading rate U, the global production rate of new seafloor at ridges, conveniently given in km^2 year^{-1}. The time for seafloor spreading to renew the Earth's surface is then $t_E \cong 4\pi r_E^2/U$. This time represents the age of old seafloor at the time of subduction; we use the subduction age and the renewal time interchangeably in dimensional arguments. The present global spreading rate is $\sim$3 km^2 year^{-1}, which would resurface the ocean basins in $\sim$100 my and the Earth in 170 my. White et al. (1992) gave a more precise estimate of 3.3 km^2 year^{-1}.

The global average heat flow depends on the square root of the renewal time:

$$q = \frac{2k\Delta T}{\sqrt{\pi\kappa t_E}} \qquad [2]$$

where ΔT is the temperature contrast between the surface and the hot mantle beneath the lithosphere, k is thermal conductivity, ρ is density, and $\kappa \equiv k/\rho C$ is thermal diffusivity. The current average heat flow in the ocean basins is $\sim$0.10 W m^{-2}, comparable to radioactive heat generation in the Archean. Radioactivity alone thus does not mandate a very vigorous form of Archean convection.

Other simple relationships bear on geodynamics. Plate thickness of the oldest lithosphere scales as

$$L = \sqrt{\kappa t_E} \qquad [3]$$

For quick calibration, 1 my lithosphere is $\sim$10 km thick. The elevation of buoyant ridges relative to old oceanic crust produces 'ridge push,' a driving force for seafloor spreading. The membrane stress (the difference between horizontal stress and the vertical stress from lithostatic pressure) scales as

$$\tau = \rho g \alpha \Delta T (\kappa t_E)^{1/2} \qquad [4]$$

where g is the acceleration of gravity and α is the volume thermal expansion coefficient. Stress is concentrated within the cool brittle upper part of the lithosphere. The integral of this stress over the thickness of the lithosphere is the 'ridge push' force available to drive plate motion. It is

$$F = \tau L = \rho g \alpha \Delta T (\kappa t_E) \qquad [5]$$

The negative buoyancy of downgoing slabs is the other major driver of plate motions.

9.06.2.3 Parameterized Convection

Equation [1] can be integrated forward or back in time if one has the surface heat flow as a function of the interior temperature (Figures 1 and 3) and the concentrations of heat-producing elements (Figure 2). From eqn [2], knowing the rate of convection as a function of interior temperature suffices. To do this, we assume that the vigor of convection depends only on the properties of the lithosphere and the uppermost asthenosphere, not those of the deep mantle. In plate tectonics terms, slabs once started down continue down into the deep mantle. In general, the global convective heat flow depends on the temperature at the base of the lithosphere and the fraction of the Earth's surface covered by the continental crust (Lenardic et al., 2005). The convective heat flow does not depend on the instantaneous rate of radioactive heat generation.

Fluid dynamists call these assumptions 'boundary layer theory.' It applies in general if the thermal boundary layers comprise a small fraction of the total thickness of the convecting region. The Earth meets this criterion; the lithosphere is much thinner than the mantle. The interaction of slabs that start from the top boundary layer and plumes that start from below is thus inevitable. As a caveat not explicitly addressed in this chapter, putative strong effects of the phase change at the base of the upper mantle and of chemical stratification within the mantle (e.g., thick dense 'dregs' in the D″ layer) violate the boundary layer assumption.

We briefly review rheology to put the remaining discussion in context. In full tensor form, the strain rate in an isotropic material is

$$\varepsilon'_{ij} = \frac{\tau_{ij}\tau^{n-1}}{2\eta\tau_{\text{ref}}^{n-1}} \qquad [6]$$

where η is the viscosity, τ_{ij} is the deviatoric stress tensor (the stress that produces shear, mathematically the stress tensor with zero trace), the scalar stress τ is the second invariant of the stress tensor $\sqrt{\tau_{ij}\tau_{ij}}$ (appropriately normalized it is the resolved shear stress in two dimensions), n is the exponent of a power-law rheology, and τ_{ref} is a reference stress that may be chosen for convenience ($1/\eta\tau_{\text{ref}}^{n-1}$ is effectively one multiplicative physical parameter that may be constrained by experiment). The scalar form is useful for dimensional calculations:

$$\varepsilon' = \frac{\tau^n}{\eta\tau_{\text{ref}}^{n-1}} \qquad [7]$$

The familiar linear or Newtonian fluid is the case where $n = 1$. A plastic substance fails at a yield stress. This corresponds to the limit $n \to \infty$. We present applications of parameterized convection in the order that their relevancies arose in the Earth's history.

A linearly viscous fluid reasonably represents liquid magma and hence the liquid magma ocean on the early Earth. It may represent convecting mush and the present asthenosphere. The formalism for this rheology illustrates key aspects of the parameterized convection approach before presentation of nonlinear $n > 1$ rheology. The convective heat flow is then

$$q = Ak\Delta T\left[\frac{\rho C g \alpha \Delta T}{\kappa\eta}\right]^{1/3} \qquad [8]$$

The dimensionless multiplicative constant A is of the order of 1; it depends on the upper boundary condition. ΔT is the temperature contrast that actually drives convection and η is

the viscosity of the convecting interior (Davaille and Jaupart, 1993a,b, 1994; Solomatov, 1995; Solomatov and Moresi, 2000). These two quantities are the dominant terms in eqn [8]. Temperature contrast is raised to the 4/3 power. Viscosity decreases orders of magnitude with increasing temperature.

The linear rheology assumed in eqn [8] may not apply to plate tectonics because trench and transform boundaries fail in friction rather than viscous flow. One approach is to apply the parameterized formula for a nonlinear fluid:

$$q = A_n k \Delta T \left[\frac{\Delta T^n (\rho g \alpha)^n}{\kappa \tau_{\text{ref}}^{n-1} \eta} \right]^{1/(n+2)} \quad [9]$$

where the full temperature contrast ΔT across the lithosphere drives convection. The dimensionless multiplicative constant A_n depends on both the exponent n and the boundary conditions. As one approaches the limit of a plastic rheology $n \to \infty$, the heat flow becomes independent of the viscosity η and proportional to square of the temperature contrast. This inference assumes correctly that the yield stress does not depend greatly on temperature; the 'brittle' strength of plates is concentrated at cool shallow depths (e.g., Zoback and Townend, 2001). Plate-bounding faults, such as the San Andreas, are much weaker than 'normal' rocks failing in friction at hydrostatic pore pressure.

We derive this temperature-squared relationship by letting τ in eqn [4] represent the plastic yield stress τ_Y, where

$$\tau_Y = \rho g \alpha \Delta T (\kappa t_E)^{1/2} \quad [10]$$

Solving eqns [2] and [10] by eliminating the resurfacing time t_E yields the heat flow

$$q = \frac{2k\rho\alpha g \Delta T^2}{\tau_Y \sqrt{\pi}} \quad [11]$$

As discussed in Section 9.06.3, the Earth's interior cooled only a few 100 K since a solid lithosphere formed. For example, the heat flow in eqn [11] when the interior was 1850 °C versus 1350 °C at present was only a factor of $(1850/1350)^2 = 1.9$ greater than present.

Stagnant-lid convection occurs beneath the lithosphere because viscosity is strongly temperature-dependent (Figure 4). As discussed in Section 9.06.5, it will become the mode of convection within the Earth in the geologic future. It is now the mode of convection within Mars and Venus. The cooler upper part of the lithosphere is essentially rigid on those planets (see Chapter 10.13).

Only a thin rheological boundary layer actually partakes in convection (Figure 4). The temperature contrast across the layer determines whether underlying convection impinges on the base of the crust. In laboratory and numerical experiments, the viscosity contrast between the top of the rheological boundary layer and the underlying half-space is a factor ~10 for a linear fluid. Simple convenient expressions arise when we represent viscosity in a traditional form:

$$\eta = \eta_{\text{ref}} \exp \left[\frac{T_{\text{ref}} - T}{T_\eta} \right] \quad [12]$$

where η_{ref} is the reference viscosity at temperature T_{ref} and T_η is the temperature scale for viscosity. The parameterized convection equation [9] in general becomes

$$q = A_n k T_\eta \left[\frac{T_\eta^n (\rho g \alpha)^n}{\kappa \tau_{\text{ref}}^{n-1} \eta} \right]^{1/(n+2)} \quad [13]$$

where the viscosity is that of the underlying half-space and the reference stress may be defined for convenience. The temperature contrast of the rheological boundary layer is

$$\Delta T_{\text{rheo}} \approx 1.2(n+1) T_\eta \quad [14]$$

Stagnant-lid convection today supplies modest amounts of heat flow to the base of oceanic lithosphere. It supplies the bulk of the heat flow through stable continental lithosphere. The asthenosphere and the rheological boundary layer beneath these regions may well behave as a linearly viscous fluid. The

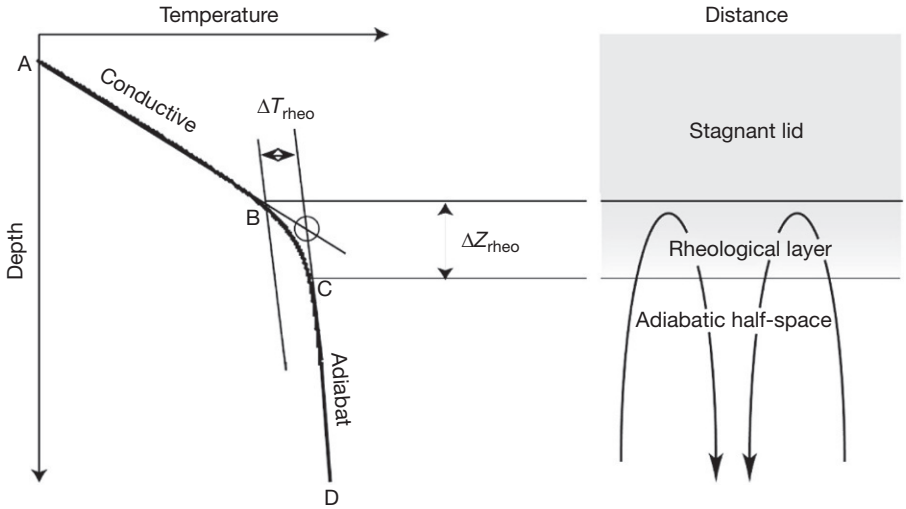


Figure 4 Schematic diagram of stagnant-lid convection at the base of the lithosphere. The geotherm (dark line, left side) has a linear conductive gradient (A to B) within the stagnant lid. It is adiabatic (C to D) below the stagnant lid. The rheological boundary layer is the top of convection (right). The geotherm changes form conductive at B to adiabat at C. The geothermal gradient in the region scales as ~½ the conductive gradient. The high viscosity near the top of the rheological layer limits the flow rate.

temperature scale T_η is not precisely known; 43 K (which provides an order of magnitude over 100 K) is a venerable approximation. It could be as high as 100 K. The temperature contrast of the rheological boundary layer is less than a few hundred K. Stagnant-lid convection does not impinge on the base of the crust within stable parts of the continent.

A caveat arises evaluating the parameterized convection equation [13] if the composition of the Earth's mantle changed over time. The thermodynamic value of T_η applies to the value of heat flow at a given time. The half-space viscosity η evolved with time as the temperature and composition of the mantle changed. The long-term effective value of T_η includes both effects. Net subduction of water could have decreased viscosity over time partly and offset the increase in viscosity from cooling (e.g., Korenaga, 2011).

9.06.2.4 Changes in the Mode of Convection

The mode of convection in the Earth changed a few times. Changes included mush-filled magma ocean to plate tectonics. (The traditional term 'magma ocean' misleadingly implies a fully molten reservoir. The nonstandard term 'mush ocean' is preferable as it implies a region of mostly crystalline mush.) Convection may have conceivably gone from plate tectonics back to mush ocean. Transition from plate tectonics to stagnant-lid convection is inevitable in the geologic future. The initiation of subduction and the delamination of the base of the lithosphere are local mode changes. ('Delamination' is massive foundering of cold dense mantle lithosphere into the underlying mantle.)

This chapter illustrates a general formalism for representing mode changes involving stagnant-lid convection. The physical processes involve the fact that the strain rate within the lithosphere depends nonlinearly on stress. We do not have a quantitative form for transitions involving a mush ocean (see **Appendix B**).

To derive eqn [14], Solomatov and Moresi (2000) presumed that stagnant-lid convection self-organizes the rheological temperature contrast ΔT_{rheo} so that flow transports the maximum convective heat flow. We illustrate this property qualitatively before doing mathematics. The convective heat flow depends on the product of the flow rate and the rheological temperature contrast. If the rheological temperature contrast is very small, little heat gets transferred because no temperature variations are available to transfer heat and the buoyancy forces driving convection are quite small. Conversely, there are ample buoyancy forces and temperature variations if the rheological temperature contrast is very large, but the high viscosity of the cool material in the stagnant lid precludes flow. The maximum convective heat flow exists in between where there are both temperature variations and a viscosity only modestly higher than that of the underlying half-space. Laboratory experiments and numerical calculations show that the following formalism provides reasonable predictions of the rheological temperature contrast.

We continue with mathematics, extending the derivation by Sleep (2002) to a more complicated rheology. To keep equations as compact as possible, we assume that linear and nonlinear creep mechanisms act in parallel and use the scalar form [7] to obtain dimensional results. The scalar strain rate is then

$$\varepsilon' = \frac{\tau}{\eta} + a\frac{\tau^n}{\eta\tau_{\text{ref}}^{n-1}} \qquad [15]$$

where a is a dimensionless constant. In general, the temperature dependence of the two creep mechanisms may differ and the stress–strain rate relationship may be more complex.

We make the rheological temperature contrast ΔT_{rheo} a free parameter with the intent of outlining the derivation of eqn [14] by finding its maxima in the convective heat flow (**Figure 5**). The thickness of the rheologically active layer is then (**Figure 4**)

$$\Delta Z_{\text{rheo}} = \Delta T_{\text{rheo}}\left(\frac{\partial T}{\partial Z}\right)^{-1} \qquad [16]$$

where $\partial T/\partial Z$ is the thermal gradient. The stress from lateral variations of temperature scales with the product of the density contrast within the rheological boundary layer and its thickness. The stress within the rheological layer is dimensionally

$$\tau \approx \rho g \alpha \Delta T_{\text{rheo}}^2 \left[\frac{\partial T}{\partial Z}\right]^{-1} \qquad [17]$$

Note that eqn [4] arises from the analogous product of density contrast with lithosphere thickness.

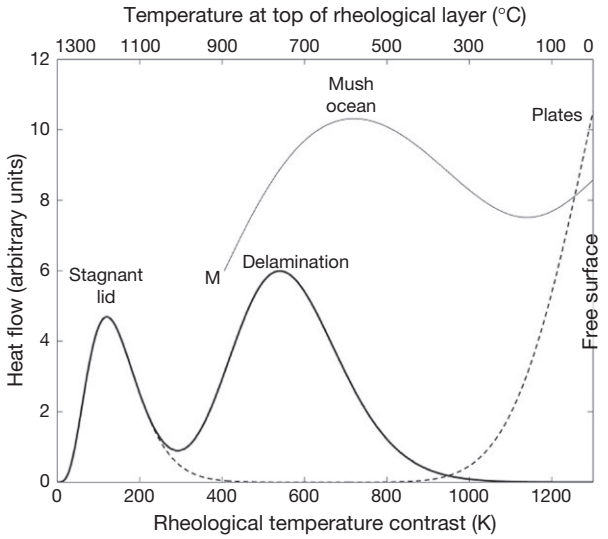


Figure 5 The heat flow in arbitrary units from eqn [19]. I have adjusted the constant a in each model so that the amplitude of the peaks is comparable for schematic plotting convenience. The second peak of the solid curve corresponds to delamination of the lithosphere, $n = 8$. The dashed curve represents plate tectonics. The computed heat flow is still increasing when the rheological temperature contrast is equal to its maximum possible value, the difference between the asthenosphere and the free surface. This defines a maximum. I represent a plastic rheology with $n = 24$. $T_\eta = 50$ K. The (top) scale for temperature at the top of the rheological boundary layer applies to the present Earth. The mush ocean curve (thin solid line) is purely schematic and applies when the interior of the Earth was much hotter than today. I draw the curve with weak dependence of heat flow on rheological temperature contrast. No stagnant-lid branch exists; the temperature contrast cannot be lower than that between upwelling mantle and nearly frozen mush (point M).

The velocity in the thermal boundary layer is the integral of the strain rate across the rheologically active boundary layer thickness, ΔZ_{rheo}, dimensionally:

$$V = \varepsilon' \Delta T_{\text{rheo}} \left[\frac{\partial T}{\partial Z}\right]^{-1}$$

$$= T_{\text{rheo}} \frac{1}{T'\eta}\left[\left(\frac{\rho g \alpha \Delta T_{\text{rheo}}^2}{T'}\right) + a\left(\frac{\left(\rho g \alpha \Delta T_{\text{rheo}}^2\right)^n}{T'^n \tau_{\text{ref}}^{n-1}}\right)\right] \quad [18]$$

where $T' \equiv \partial T/\partial Z$ compacts notation. The convective heat flow is dimensionally $\rho C V \Delta T_{\text{rheo}}$. Finally, the uppermost part of the rheologically active boundary layer must deform for flow to occur. The effective value of viscosity-resisting flow is that where the temperature is about ΔT_{rheo} below that in the underlying half-space. Applying the viscosity equation [15] yields an equation for the convective heat flow

$$q = \rho C \varepsilon' \Delta T_{\text{rheo}}^2 \left[\frac{\partial T}{\partial Z}\right]^{-1}$$

$$= \Delta T_{\text{rheo}}^2 \frac{\rho C}{T'\eta_0}\left[\left(\frac{\rho g \alpha \Delta T_{\text{rheo}}^2}{T'}\right) + a\left(\frac{\left(\rho g \alpha \Delta T_{\text{rheo}}^2\right)^n}{T'^n \tau_{\text{ref}}^{n-1}}\right)\right] \exp\left[\frac{-\Delta T_{\text{rheo}}}{T_\eta}\right] \quad [19]$$

where η_0 is the viscosity of the underlying half-space. The expression is at a maximum when $\partial q/\partial T_{\text{rheo}} = 0$. One recovers results for power-law rheology by retaining only the a term in eqn [19]. One recovers the $n+1$ dependence in eqn [14]; the factor 1.2 is obtained from experiments. One obtains eqn [13] with the $1/(n+2)$ exponent by setting the convective heat flow in eqn [19] equal to the conductive heat flow $q = k\partial T/\partial Z$ and solving for the temperature gradient.

For application to the Earth, two maxima occur in eqn [19] if the first term dominates at small stresses and small values of the rheological temperature contrast (Figure 5). The first maximum represents stagnant-lid convection within a fluid with linear rheology. The second high-stress maximum might represent other situations.

Figure 5 shows a case where the second maximum is within the lithosphere. This represents conditions for delamination of the base of the lithosphere. Ordinarily, the stagnant-lid mode is stable. From eqn [17], the stresses associated with its small rheological temperature contrast ΔT_{rheo} are much less than those of the delamination mode. As the rate of nonlinear creep depends on the rheological temperature contrast raised to a high power, the upper part of the lithosphere stays rigid.

Delamination requires a large stress perturbation scaling to the stress implied by its ΔT_{rheo} in eqn [17]. In the Earth, this stress could arise from the impingement of a mantle plume, continental collision, low-angle subduction, or continental breakup.

The upper boundary condition on the lithospheric mantle affects delamination. In particular, the hot base of overthickened continental crust acts as a free surface (Appendix B). In terms of the heat flow equation [13], the dimensionless constant A_n is larger for a free surface boundary than a rigid boundary. In terms of eqn [18], material flow velocity is higher if the upper boundary slips freely than if this boundary is rigid. With regard to Figure 5, the presence of fluid continental crust affects delamination if the temperature at its base is within the delamination maximum.

Plates are rheological boundary layers extending all the way to the free surface of the Earth. The rheological temperature contrast is thus the full temperature contrast across the lithosphere. On the Earth, existing plate tectonics maintains the high stresses needed to nucleate new plate boundaries as old ones are consumed by continental and ridge–trench collisions. The plate tectonic maximum in Figure 5 is the intersection of the heat flow curve with the temperature of the free surface. The author's statements about the transition of stagnant-lid convection to delamination also apply to its transition to plates. Stagnant-lid convection may not produce large stresses in a one-plate planet, making transition to plate tectonics difficult. There is little evidence of it occurring in the geologic records of the planets we have observed. Silver and Behn (2008) suggested a global hiatus between 1 and 1.75 Ga in plate motions based on the occurrence of passive margins. The difficulty of restarting subduction does not apply if plate movements were merely sluggish over that interval. Mechanically, Solomatov (2004) showed with dimensional and numerical models that this transition may occur under realistic situations.

The transition of mush ocean to plate tectonics is also relevant to the Earth (Appendix B). Figure 5 illustrates a time when the Earth's interior was hot enough to maintain a mush ocean. As drawn, maxima correspond to mush ocean convection where the base of basaltic crust founders and plate motion where the entire crust founders. The rheological temperature contrast is greater than the contrast between solid upwelling mantle and nearly frozen mush. The heat flow is weakly dependent on the rheological temperature contrast. The mush ocean may have behaved as a mixture of these modes.

9.06.2.5 Pressure-Release Melting at Mid-Ocean Ridges

Upwelling mantle melts beneath mid-ocean ridges. The melt ascends and freezes to form the basaltic oceanic crust. Both the basaltic crust and the depleted residual mantle are less dense than the melt source region from which they differentiated. These density changes are large enough to affect plate dynamics. Lithosphere formed from hot mantle is hard to subduct (e.g., Davies, 1992; Herzberg et al., 2010; Sleep and Windley, 1982; van Thienen et al., 2004).

Following Klein and Langmuir (1987), McKenzie and Bickle (1988), Plank and Langmuir (1992), Klein (2003), and Herzberg et al. (2010), we present a short discussion of how this process affects plate motion. Crudely, the upper Earth's mantle consists of olivine, orthopyroxene, clinopyroxene, and an alumina-bearing phase. Garnet, a dense mineral, bears the Al_2O_3 at mantle depths below ~100 km. At crustal depths, low-density feldspar contains most of the Al_2O_3. Mid-ocean ridge basalt that has a near-surface density of ~3000 kg m^{-3} is eclogite below 150 km depth at the interior temperature of the mantle with a density of 3600 kg m^{-3} at current plume temperatures (Sobolev et al., 2005). The stability field of eclogite is temperature-dependent; subducted basalt begins to form eclogite at ~30 km depth. The density difference of basalt with respect to eclogite is significant. For comparison, the volume thermal contraction coefficient of mantle is 3×10^{-5} K^{-1}; a temperature change of 1000 K produces a density change of only ~100 kg m^{-3}.

Melting occurs as pressure decreases within the mantle upwelling beneath ridges. Eutectic melting illustrates the mass balance at low melt fractions although it is an inadequate representation of melt chemistry. In more detail, melts segregate from the ascending mantle as they form and then pool into the magma chamber of the ridge axis (e.g., Herzberg and Rudnick, 2012). Clinopyroxene and the alumina-bearing phase are exhausted before olivine and orthopyroxene; further temperature increases result in only modest amounts of additional melting of this residuum.

For additional simplicity, we let the melting temperature be a linear function of depth:

$$T_m = T_0 + \beta Z \quad [20]$$

where T_0 is the melting temperature at the surface and β is the gradient of melting temperature with depth Z. The temperature of the upwelling material in the absence of melting is

$$T_a = T_p + \gamma Z \quad [21]$$

where T_p is the potential temperature and γ is the adiabatic gradient. Melting begins at the depth where the curves intersect

$$Z_m = \frac{T_p - T_0}{\beta - \gamma} \quad [22]$$

The temperature above this depth lies on the melting curve. The specific heat for cooling the material from the solid adiabat equals the latent heat for partial melting

$$\rho C(T_a - T_m) = LM \quad [23]$$

where L is the latent heat per volume and M is the fraction of partial melt. Solving for the fraction of melt gives

$$M = \frac{\rho C}{L}\left[T_p - T_0 - (\beta - \gamma)Z\right] \quad [24]$$

The total thickness of melt production is crudely that in a column that reaches the surface

$$\Phi = \int_0^{Z_m} M\,dZ = \frac{\rho C(T_p - T_0)^2}{2L(\beta - \gamma)} \quad [25]$$

The actual mantle is not precisely eutectic. The heat balance equation becomes

$$\rho C(T_a - T_m(M, Z)) = LM \quad [26]$$

where the melting temperature is a function of depth and fraction of melt. One solves eqn [26] numerically and integrates the fraction of melt upward (as $M/(1-M)$ to correct geometrically for compaction of the matrix). For a ridge axis, the computed thickness of melt equals the depth to the base of the crust. One truncates the integral at that depth. One truncates at the base of the lithosphere for midplate melting.

It is convenient to calibrate the eutectic model for quick calculations so that it represents melt fractions for the actual melting in the present Earth, so it is unnecessary to precisely know the actual absolute temperature of ambient mantle. Here, $(\beta - \gamma)$ is 3.2 K km^{-1} and $L/\rho C$ is about 840 K. The modern depth of extensive melting Z_m is 56 km. These parameters give 6 km of average oceanic crust. The temperature change per thickness of crust is

$$\frac{\partial T_p}{\partial \Phi} = \frac{L(\beta - \gamma)}{\rho C(T_p - T_0)} = \frac{L}{\rho C Z_m} \quad [27]$$

This expression yields the expected value of 15 K km^{-1}.

These considerations imply that oceanic crust was much thicker in the past when the interior of the Earth was hotter (Figure 6). Consider a time in the past when the mantle was 200 K hotter than now. The oceanic crust was ~14 km thicker than now or 20 km thick. It everywhere resembled the thick hotspot crust beneath Iceland. This buoyant crust did not easily subduct until cooling had progressed downward into the mantle lithosphere. Present-day oceanic crust, such as remnants of the Farallon Plate (Atwater, 1970; Wilson et al., 2005), became difficult to subduct at 5–10 my age. As cooling time depends on the square of the depth, a 20 km thick crust would be difficult to subduct at $(20/6)^2 = 11$ times these ages or 55–110 my.

Plate tectonics were even less efficient in subducting thicker crust when the mantle was hotter than in the example. This implies that convection within the partially molten mush carried heat to the surface. The next section considers this epoch.

Davies (2006) suggested an alternative where subduction remained efficient in the Archean. Dense subducted crust separated from slabs as they started to warm up. This basaltic component accumulated in the deep mantle. The remaining mantle that upwelled at ridge axes was depleted in the basaltic component. Less crust formed than at present for a given temperature. The lower mantle may have even episodically

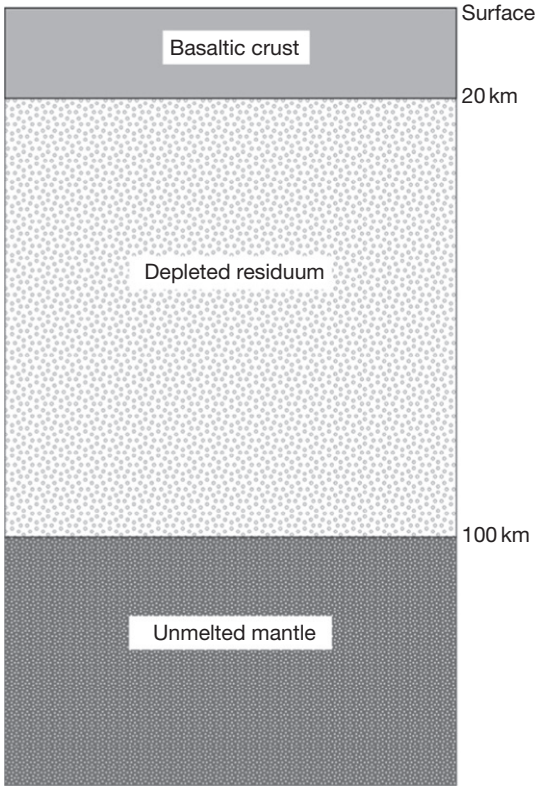


Figure 6 Section through oceanic lithosphere where upwelling mantle starts melting at 100 km depth. This lithosphere is difficult to subduct. Both the crust and depleted residuum are buoyant relative to the underlying mantle. Unlike basaltic mush, the depleted residuum is quite viscous and does not internally convect.

convected separately from the upper mantle (Davies, 2008). However, Herzberg et al. (2010) noted that putative examples of Archean nonarc mafic rocks chemically resemble those from modern oceanic plateaus where hot plumes feed ridge axes. Modern oceanic plateaus have thick crust, which is difficult to subduct. For reference, basaltic and peridotite domains do not appear to separate nor chemically homogenize within the modern deep mantle. Rather, the postperovskite transition behaves as a mechanical mixture of these materials (Grocholski et al., 2012).

In general, hypotheses for rapid ancient plate motions including those with thin Archean oceanic crust require that the Urey ratio of radioactive heat production to mantle heat flow be relatively high ~0.70 (Herzberg et al., 2010). Otherwise, models with low radioactivity cool rapidly with time implying excessive temperature in the past. Hence, these models imply sluggish ancient plate motions.

9.06.3 Aftermath of the Moon-Forming Impact

The violent birth of the present Earth makes for simple physics (Sleep et al., 2001; see also **Chapters 9.04** and **9.01**). A Mars-sized body collided with a Venus-sized body at ~4.5 Ga, leaving the Earth–Moon system. The core of the projectile merged with the target's core.

Tidal dissipation from the Moon likely prolonged the cooling time of the Earth to several million years; the present lunar orbit indicates that it moved out slowly from the Earth over this time. We deduce a trial early sequence of events without having to precisely know physical properties following Zahnle et al. (2007).

A gas–liquid surface formed quickly. The critical point of silicates is approximately that of SiO_2, ~13 500 K, 6.4 kbar (0.64 GPa) (Ahrens and O'Keefe, 1972). Any material that was hot rose to the top of the atmosphere and cooled very quickly by radiation to space. There was not enough energy to raise the entire mantle onto an adiabat with this condition. Once this very hot material had cooled, the mantle adiabat allowed partial vaporization of silicates well above ~3350 K (the boiling point of olivine at 1 bar from Ahrens and O'Keefe, 1972). The rock vapor radiated heat, quickly forming silicate cloud tops at ~2300 K. The photosphere radiated at this temperature for less than 2000 years. During this time, all mantle material circulated repeatedly through the cloud tops.

Once the surface temperature dropped below ~3350 K, the liquid was to the first order of bulk mantle composition, and more volatile minor mantle components entered the gas. The partition of components between liquid mantle and gas has not been measured at this temperature. Basalt is a reasonable analog and volatile solubility at low pressures weakly depends on temperature (Hirschmann, 2012). Carbon dioxide is nearly insoluble in magma at low pressures, and its bulk inventory entered the atmosphere, giving ~100 bars partial pressure. Much of the Earth's inventory of water entered the atmosphere. The total pressure was thus similar to that at the modern seafloor.

A significant concentration of alkali silicates, NaCl, KCl, and sulfur species also entered the gas. Volatiles in seafloor magma and hydrothermal systems provide guidance on these components, especially after a solid scum had formed. The silicate vapor components were alkali-rich; vapor formed from continental crust in asteroid impacts and the resulting tektites are proxies. The vapor pressure of tektites extrapolates to ~3 bars at 3350 K and 0.1 bars at 1640 K (Walter and Carron, 1964). Volatiles in seafloor komatiite flows are another proxy (Stone et al., 1996). Schaefer et al. (2012) modeled the thermodynamic behavior of atmospheres above very hot extrasolar planets, but not the Moon-forming impact in any detail.

Silicate vapor condensed with increasing elevation above the magma ocean along gas adiabats forming opaque fog, rain, and virga. NaCl formed clouds at higher elevations; the partial pressure of NaCl is ~0.01 bar at 1275 K (Kvande et al., 1979). Sulfur is complicated as the condensing species are not obvious; elemental sulfur boils at 1 bar and 718 K. Overall, (1) the atmosphere was opaque up to its highest thick cloud top. (2) The appropriate adiabat was 'wet' where bar quantities of silicate, NaCl, KCl, and sulfur species condensed. (3) Water clouds formed at some time, but calculations are required to determine the surface temperature at which this occurred. Hot radiation from the volcanic surface would not reach the uppermost cloud deck, as lower cloud layers were opaque. In particular, the sodium chloride cloud top was cooler than 1500 K, the classical surface radiating temperature below which water vapor is opaque. (4) Complexing by sulfur and chlorine (as in ore-forming fluids) may have caused some trace elements (perhaps W and Pt) to behave as volatiles. Chemical and isotopic fractionation from these processes may persist on the Earth, but is not considered further in this chapter.

Dissipation of lunar tides became important once the deep mantle began to solidify. The surface temperature was then about 2800 K (Andrault et al., 2011, Figure 5). (Liquid adiabat and the liquid in the lower mantle might be such that freezing started in the middle of the lower mantle, but that does not affect the basic physics presented here.) Tidal dissipation was unimportant before solids formed, as molten silicate and iron have very low viscosities. The solid material had a viscosity appropriate for dissipation, but heating from dissipation remelted it, acting as a buffer on mantle temperature that allowed only a modest thickness of solid to form. Tides dissipated heat only at the rate that it could escape from the top of the atmosphere. Water clouds at the top of the atmosphere were likely by this time, so that the planet radiated near the runaway greenhouse threshold of 150 W m^{-2}. It took a few million years for the Moon to move far enough away that tidal dissipation could no longer supply this heat flow. In a few more million years, the mantle froze with the remaining liquid reaching an adiabat where this solid rind formed at the surface. The mantle was solid below ~1175 km when this happened (Andrault et al., 2011, Figure 5).

The upper mantle then froze to mostly solid much. Solid-state convection could not maintain a runaway greenhouse. Water is weakly soluble in uppermost mantle solids near the solidus. Thus, most of the Earth's water, CO_2, and NaCl (or KCl) entered in the atmosphere. NaCl rained out when the surface was 1500–1250 K. The NaCl-rich fluid gradually became enriched in water until the critical point of ~670 K was reached. Oceans then covered the Earth.

The Earth lost far more heat during these molten epochs than during its subsequent history. At the time the magma

ocean froze to mush, it was only a few 100 K hotter than the modern mantle. This fossil heat and radioactivity drove tectonics for the subsequent ~4.5 billion years. Chemically dense dregs may well have accumulated at the base of the mantle over this interval, thereafter thermally insulating the core from the mantle. This process is important with regard to mantle plumes.

Thereafter, sunlight and the concentration of greenhouse gases controlled climate with liquid water at the surface. An uninhabitable ~200 °C surface greenhouse likely persisted until carbonate formation and crustal overturn sequestered the CO_2 in the subsurface.

9.06.3.1 Early Mush Ocean

The liquid magma ocean ended about the same time that the surface froze to a kilometer-scale lid (Sleep et al., 2001; Zahnle et al., 2007; also see **Chapters 9.04** and **9.01**). Modern fast ridge axes provide analogy to the subsequent mush ocean. Magma entering the axis freezes quickly (Sinton and Detrick, 1992). The bulk of the 'magma' chamber is mostly crystalline mush. Only a thin (tens of meters) lens of fully molten rock exists at the axis.

These observations show that molten mafic bodies quickly freeze to mush. They provide quantification on the rate that liquid convection within the chamber delivers heat to the surface. The parameterized convection equation [8] for a linear rheology of basaltic melt applies. Current ridge axes provide a minimum estimate for the convective heat flow from the liquid magma ocean. Molten material of mantle composition is orders of magnitude less viscous than modern basalt (Liebske et al., 2005).

Estimating the heat flow through a modern magma lens at the ridge axes driven by latent heat is straightforward. We balance the heat per time per length of ridge with the heat flow per length of ridge axis out of the width of the magma lens. This yields that

$$q = \frac{LZ_c v}{W} \qquad [28]$$

where $L = \sim 1.5 \times 10^9$ J m^{-3} is the latent heat of freezing, W is the width of the magma lens, Z_c is magma chamber thickness, and v is the full spreading rate (Sleep et al., 2001). The fastest spreading centers have the highest heat flow. We retain parameters from the work of Sleep et al. (2001). The highest full spreading rate on the modern Earth is 155 mm year^{-1}, magma chamber thickness is ~5 km, and the lens width is 1 km. These quantities yield a heat flow of 40 W m^{-2}. Fontaine et al. (2011) obtained 30 W m^{-2} with numerical models using this approach. The latent heat of the entire mantle could maintain these heat flows globally for only 3–4 my.

This reasoning implies that any fully molten magma ocean was moderately short-lived on the Earth. That is, cumulate igneous rocks formed at mid-lower mantle depths only in the immediate aftermath of the Moon-forming impact. The phase assemblages of the lower mantle and their partitioning of trace elements between melt and cumulate solids provide means of recognizing later melts derived from these cumulates.

Rizo et al. (2011) examined 3.7 Ga basaltic rocks from the Isua region of Greenland in this regard. Their combined use of the systems ^{146}Sm–^{142}Nd, ^{147}Sm/^{143}Nd, and ^{176}Lu/^{176}Hf constrains the lower mantle chemical origin of the source reservoir and its age. In particular, the half-life of ^{146}Sm is ~68 my compared with the older value of ~103 my (Kinoshita et al., 2012); only processes on the early Hadean Earth could have produced variations in ^{142}Nd/^{144}Nd. In general, studies of short-lived radioactive decay systems yield computed ages referenced to the formation age of the solar system obtained from long-lived isotopic systems.

The cumulate model ages cited by Rizo et al. (2011) need to be adjusted for the revised value of the ^{146}Sm half-life as well as for their recomputed initial isotopic ratio of the Earth (Kinoshita et al., 2012). These ages are also dependent on the assumed initial bulk Sm/Nd composition of the mantle. Rizo et al.'s (2011) model ages of 4.53 Ga for ordinary chondrite composition and 4.47 Ga for enstatite-chondrite composition do not need to be adjusted significantly (Kinoshita et al., 2012, Table 1). Their computed age for Sm/Nd above chondrites (4.37 Ga) needs to be increased to ~4.42 Ga (Kinoshita et al., 2012, Table 1). These ages are old enough that the presence of hot material that could be molten and crystallize cumulates at lower mantle conditions is reasonable.

The younger of the model ages of 4.42 Ga most likely applies and bears on geodynamics; it is a calibration point for thermal modeling and involves chemical tracer. The ^{142}Nd/^{144}Nd ratio of modern ridge and ocean island basalts is homogeneous within resolution and consistent with the mantle with Sm/Nd 6% greater than chondrites (Jackson and Carlson, 2012). A complimentary mantle reservoir with low ^{142}Nd/^{144}Nd has not been detected. One would not exist if the mixture of material that accreted into the Earth had high Sm/Nd or if complementary high Nd/Sm crustal material was returned to space by impacts early in the Earth's history. A partially 'hidden' reservoir with low ^{142}Nd/^{144}Nd could still exist; Jackson and Carlson (2012) pointed out that it could contribute up to ~14% of the Nd to some individual lavas still not be resolved. At a minimum, the ^{142}Nd/^{144}Nd data show that Hadean mantle processes were sluggish enough that some lower mantle cumulates survived until at least 3.7 Ga. Mixing was vigorous enough that the nonlithospheric mantle has not sourced igneous rocks with ^{142}Nd/^{144}Nd differing from the modern mantle value since that time.

In analogy with modern ridge axes, convection in the underlying solid mantle limited heat transfer through the mush ocean. This process has many aspects of stagnant-lid convection (**Appendix B**). The viscosity of the solid mantle changes greatly with cooling, controlling the vigor of convection. However, magma from the underlying solid convection sees the floor of the mush chamber as a nearly fluid, partly permeable interface. The multiplicative constant in eqn [9] is thus larger than the multiplicative constant for convection beneath a rigid stagnant lid. The relevant temperature difference driving convection in eqn [9] scales to the difference between the liquidus and the solidus. This quantity decreased modestly as the mush ocean cooled. The crustal lid of the mush ocean foundered and deformed, grossly like plates.

The heat balance equation [1], the convection equation [9], and the strong dependence of viscosity on temperature [12] imply quick cooling at high temperature and low viscosity followed by slower cooling (**Figure 3**). The Earth transitioned

from its earliest history where radioactive heat generation is a trivial item in the heat budget to the present status where radioactive heat generation is in crude steady state with surface heat flow.

9.06.3.2 Steady-State Mush Ocean and Transition to Plate Tectonics

As noted in Section 9.06.2.1, radioactive heat generation in the later Hadean was comparable to the modern global average heat flow ~0.07 W m^{-2}. In the models, the heat flow is ~0.13 W m^{-2} at 4.0 Ga. Global tectonic rates of $(0.13/0.7)^2 = 3.4$ times present could have vented this heat. This radioactive heat production is low enough to permit a significant hiatus in tectonic vigor at the time of transition between the mush ocean and plates. The radioactive heat generation would increase mantle temperature ~42 K in 100 my in the absence of heat flow from the core and out of the surface. For example, a total hiatus of 250 my would warm the mantle by only 100 K.

In terms of eqn [1], the heat flow and the interior temperature need not have decreased monotonically with time (Lenardic and Crowley, 2012; Sleep, 2000; Stevenson, 2003). The surface heat flow as a function of temperature may be multivalued, here with branches for plate tectonics and mush ocean (Figure 1). The mush ocean could well have mildly overcooled the mantle (Figures 2 and 3). Hot buoyant material rose to the top of the mantle, cooling near the surface of the mush ocean. The cooled material foundered, eventually filling the mantle from bottom up. The mush ocean ceased when the cooled material filled the mantle (point A to point B, Figure 1). Sluggish subsequent plate tectonics persisted for several hundred million years until radioactivity heated up the mantle enough to restart the mush ocean (point C to point D). The heat flow and interior temperature looped over points ABCD until the heat flow from plates (point B) was less than radioactive heat generation.

Herzberg et al. (2010) and Herzberg and Rudnick (2012) preferred a single increase in temperature during the Archean reaching its peak at ~3.0 Ga (Figure 3). They deduced that the peak of continent formation was between 2.5 and 3.5 Ga at the time of maximum temperatures. The model shown illustrates a moderate amount of overcooling. An extreme form of overcooling is unlikely as a dense atmosphere blanketed the Earth after the Moon-forming impact. Much of the lower mantle was already frozen but hot before freezing commenced at the surface. The surface initially was modestly below the solidus rather than clement. Greenhouse conditions at ~200 °C and perhaps clement conditions prevailed shortly after the end of the mush ocean. In any case, there was insufficient radioactivity in the mantle to reheat it more than a few hundred K.

As noted later in this subsection, the brief mush ocean episodes could be periods of rapid continent generation because hydrated crust subducts into basaltic mush. The ascending water causes the mush to melt extensively producing a granitic (*sensu lato*) magma.

An alternative possibility is that the mush ocean transitioned monotonically to sluggish plate motions inhibited by thick oceanic crust as discussed in Section 9.06.2.5 (Korenaga, 2006). Plate rates were slower than present. Mush ocean and conventional plates prevailed in different regions. The mantle temperature remained at the transition in the monotonic model (Figure 3) and the heat flow remained between the plate and mush ocean values (Figure 2). Eventually, radioactive heat production decreased to where plate tectonics could vent it. Plate tectonics thereafter cooled the Earth.

The transitions as drawn are discontinuous. Any strong dependence of global heat flow on interior temperature will trap the mode of convection in transition until radioactive heating wanes enough to be balanced by the less vigorous mode. In the early days of plate tectonic theory, it was thought that the dependence of heat flow on interior temperature was so strong that the Earth achieved its present interior temperature soon after it accreted (Tozer, 1970). This concept applies in the monotonic model over much of the early history of the Earth, though not now.

We apply Section 9.06.2.5 to put the above heat balance inferences into a petrologic context. Conditions for maintaining the mush ocean in steady state are relevant to the causes of the mush ocean's eventual demise. The mush ocean consists of a frozen crust with some continental material, melt lens where ascending magma ponds and freezes to mush, a thick layer of mush, and the underlying mantle (Figure 7). As already noted, solid-state convection in the mantle limits the rate of heat transfer (Appendix B).

The base of the mush acts as a permeable interface. Ascending mantle melts. Dikes ascend through the mush and into a magma lens. Some of the magma freezes to mush. Some freezes to form basaltic crust. The crustal volume stays in steady state; cool thick regions of the crust founder into the mush in a process like subduction. The dense base of the mush and foundered crust sink into the mantle, balancing the magmas injected from the mantle.

We focus on the requirement that the supply of melt equals the return of melt to the mantle. Both processes depend on the temperature of the upwelling mantle. If the mantle is hot enough, the thickness of the mush lens self-organizes to a steady state.

First, the frozen mush at the base of the ocean needs to be dense enough that it sinks into the mantle. Early in the Earth's history, the mantle was hot enough that upwelling material melted completely. The melt composition hence was mantle composition. Cool material was denser than hot material. Mush sank once frozen.

Later on, the mantle cooled to a temperature where basalt was the most voluminous melt. The cooled basaltic crust has about the same composition as the underlying mush. It can subduct and founder into the mush ocean. However, as already noted, basaltic material and its depleted residuum are less dense than unmelted mantle at shallow depths. However, the dense mineral garnet makes basaltic composition heavier below ~150 km depth. (Sobolev et al. (2005) discussed melting of the eclogite component in plumes.) The base of the mush ocean sinks into the mantle if it is deeper than that depth, but not if it is shallower. This feature tends to prevent the base from getting deeper than ~150 km depth.

Second, the balance between recharge from basalt from the upwelling mantle and subduction of slabs to the mantle determines whether the mush chamber thickens or thins. The models assume that the chamber (in the plate mode) thickens once the mantle becomes hotter than 1600 °C, the potential

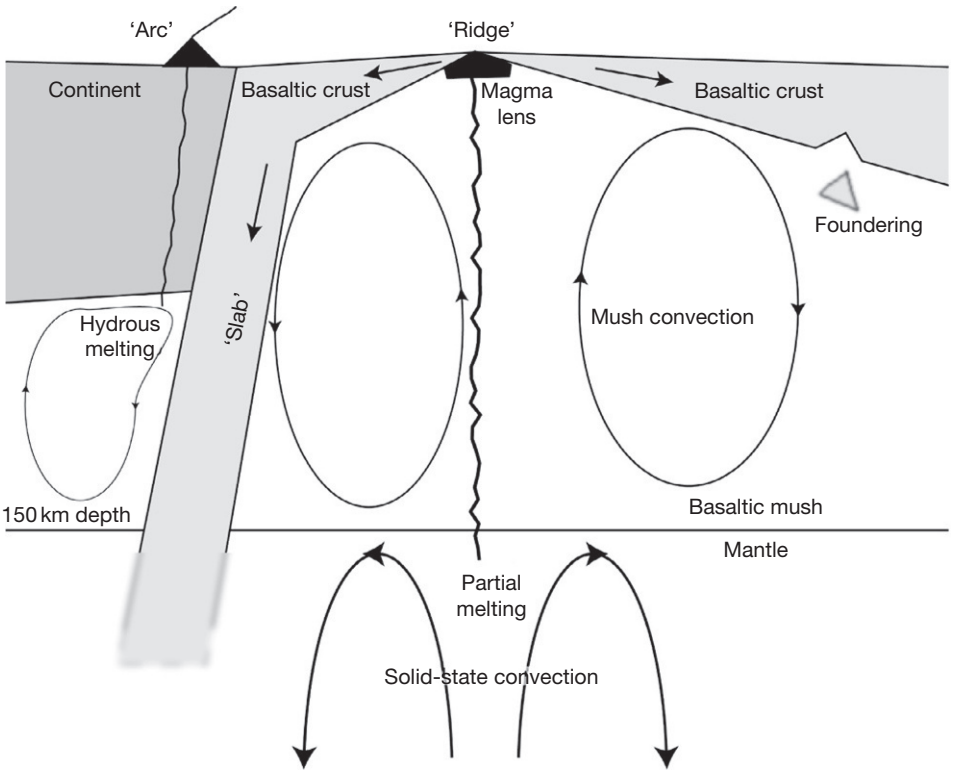


Figure 7 Schematic diagram of mush ocean with vertical exaggeration. Analogs to plate tectonic features are in quotes. Upwelling mantle partially melts. Melt ascends into the magma lens at the ridge axis and freezes, making basaltic crust and basaltic mush. Basaltic crust both subducts as slabs and founders as blocks. Partial melting of hydrated basalt produces arcs and continents. Convection within the mush transfers additional heat from the mantle. Slabs, entrained mush, and foundered blocks return basaltic material to the mantle. Elsewhere, thick lithosphere composed of depleted mantle residuum stabilized some continental regions. The figure represents a putative Hadean process as well as the net effect of a widespread but regional Archean process in areas of thickened mafic crust.

temperature where significant melting starts at 150 km depth in eqn [22]. The melt flux needs only to replenish the basaltic material that sinks back into the mantle. It does not have to be hot enough to generate a 150 km thickness of melt in eqn [25].

From this reasoning, the mush chamber becomes unstable to depletion once the mantle temperature drops below that needed for recharge. The rounded potential temperature where this happens in the models is 1550 °C, where melting in eqn [22] begins at 118 km depth. In the nonmonotonic model (**Figure 1**), convection jumps to plate tectonics. The remaining mush freezes to lithosphere and the foundering regions of crust become slabs. At this time, the ascending mantle generates only 20 km of oceanic crust in the linearized model (**Figure 6**). This is thick enough that plate tectonics cannot vent the heat generated by radioactivity (**Figure 2**). The mantle heats up until enough ascending material melts for the mush ocean to restart (**Figure 3**).

Alternatively, as in the monotonic model, convection becomes sluggish with only regional and transient regions of mush ocean, as the mantle approaches the critical temperature to maintain the mush ocean. The convective heat flux balances radioactive heat generation during a long transition period that lasts until after the end of the Archean in the monotonic model.

It is obvious that the mush ocean as envisioned provides an efficient mechanism for generating continental crust at arcs. Water from the slab enters hot gabbroic mush, rather than mantle peridotite in standard subduction. The melt is granitic (*sensu lato*). There are no data to appraise this process in the Hadean. However, this construct does serve to represent the net effect of more complicated regional Archean processes in **Section 9.06.4.4**.

9.06.3.3 Zircon Evidence

As discussed in the last section, Earth scientists know too little about the material properties of the Earth to make a more quantitative prediction of the vigor of a mush ocean or the timing of its demise. The crust of Mars may provide some analogy with future exploration. Right now, the mush ocean is the first epoch of Earth's interior with even a meager geologic record. Geochemists have extracted detrital zircon crystals from Archean and Proterozoic quartzites in Australia (e.g., Bell et al., 2011; Cavosie et al., 2004; Dunn et al., 2005; Harrison, 2009; Hopkins et al., 2010; Thern and Nelson, 2012). The igneous ages of oldest crystals are ~4.4 Ga, a time when the mush ocean might have persisted. The oldest highly precise age is 4.374 Ga (Valley et al., 2014). The old zircons resemble those from modern granites (*sensu lato*) (Cavosie et al., 2004; and references therein). In particular, some of the granites formed at the expense of sediments derived from meteoric weathering.

Isotopic studies of 4.01–4.37 Ga zircons provide evidence of an older separation of the parent element Lu from the

daughter element Hf in the Earth's interior. Within the resolution of a few million years after the Moon-forming impact, fractionated continental-type crust with elevated Lu/Hf covered part of the Earth's surface (Harrison et al., 2005). Its depleted low-Lu/Hf complement comprised a significant mantle reservoir.

These data imply some form of crustal overturn analogous to modern subduction. Hydrous mafic rocks either melted directly or dehydrated causing the surrounding hot mush to partially melt. The surface needed to be cool enough for liquid water; a warm greenhouse at ~200 °C suffices as hydrous minerals are stable at black-smoker temperatures of 350 °C. The hydrated rocks needed to founder. Getting buried by subsequent lava flows or more traditional subduction suffices. They then needed to be heated, no problem given a mush ocean. That is, the typical magmatic rocks associated with subduction are not strong indications of the modern process, especially when they are out of context in detrital zircons.

Cavosie et al. (2004) presented a potential constraint on the vigor of crustal recycling processes after 4.4 Ga. Age gaps and clusters exist within their sample suite, just like with a modern orogenic belt (**Figure 8**). Thern and Nelson (2012) examined the statistics of a much larger suite from multiple areas in the Yilgarn Craton of Australia. They resolved a 4.15–4.36 Ga suite and 3.91–4.15 suite. From the geographic distribution of samples and Archean zircon ages, they deduced that terrains bearing each suite docked at ~3.747 Ga. The few 0.1 Ga quiescent duration of the terrains before docking is comparable to modern examples and does not require particular rapid plate motions. In any case, tectonics after 4.36 Ga were sluggish enough that some Hadean continental crust survived to at least the end of the Archean.

Additional studies support this inference. Tessalina et al. (2010) reported $^{142}Nd/^{144}Nd$ variations in Archean cherts and lavas that were likely derived at the expense of Hadean rocks that persisted at that time. Greenstones within the Nuvvuagittuq region of Québec have $^{142}Nd/^{144}Nd$ variations (Adam et al., 2012; O'Neil et al., 2008, 2011, 2012). O'Neil et al. (2012) proposed that these mafic rocks erupted at ~4.4 Ga and then were intruded by gabbros at ~4.2 Ga, using the revised half-life of ^{146}Sm (Kinoshita et al., 2012). These rocks experienced later metamorphic events. At a minimum, the interpretation indicates that the mantle had cooled into the temperature range for basalt formation, but had not grossly overcooled below this range.

However, Kemp et al. (2010) and Arndt and Nisbet (2012) interpreted hafnium isotopic data on Hadean zircons to indicate that the Hadean crust continually remelted. This process may have mainly involved highly radioactive felsic crust that could generate enough heat to melt. These samples were obtained from the same localities as the Thern and Nelson (2012) study. Repeated melting is a viable mechanism to transfer heat through the Hadean crust, so are Hadean tectonics with protracted inactivity. The first situation contrasts with Archean where the base of the continental crust did not undergo continuous melting to form granites (Burke and Kidd, 1978).

The hafnium–tungsten system provides analogous information; ^{182}Hf decays to ^{182}W with a half-life of ~8.9 my. Reservoirs with variable $^{182}W/^{184}W$ formed only in the early history of the solar system. Tungsten strongly entered the Earth's core, while Hf remained with the silicate fraction. Magmatic processes also cause W/Hf to vary within silicates. Touboul et al. (2012) described 2.8 and 3.5 Ga komatiites that are enriched in ^{182}W relative to typical terrestrial rocks. The source regions of these magma escaped mixing with low $^{182}W/^{184}W$ material that contaminated the rest of the mantle during the late heavy bombardment at ~4 Ga (**Section 9.06.3.4**) or even the Moon-forming impact or formed by magmatic processes on the very early Earth. Iizuka et al. (2010) measured tungsten isotopes in the 3.6–4.1 Ga Acasta Gneiss and ~3.8 Ga rocks in Greenland that have ^{142}Nd anomalies. The $^{182}W/^{184}W$ ratio of these rocks is the typical terrestrial value. These data constrain that the source reservoir became isolated from the mantle greater that 40 my after solar system formation when ^{182}Hf was extinct, but while ^{146}Sm persisted.

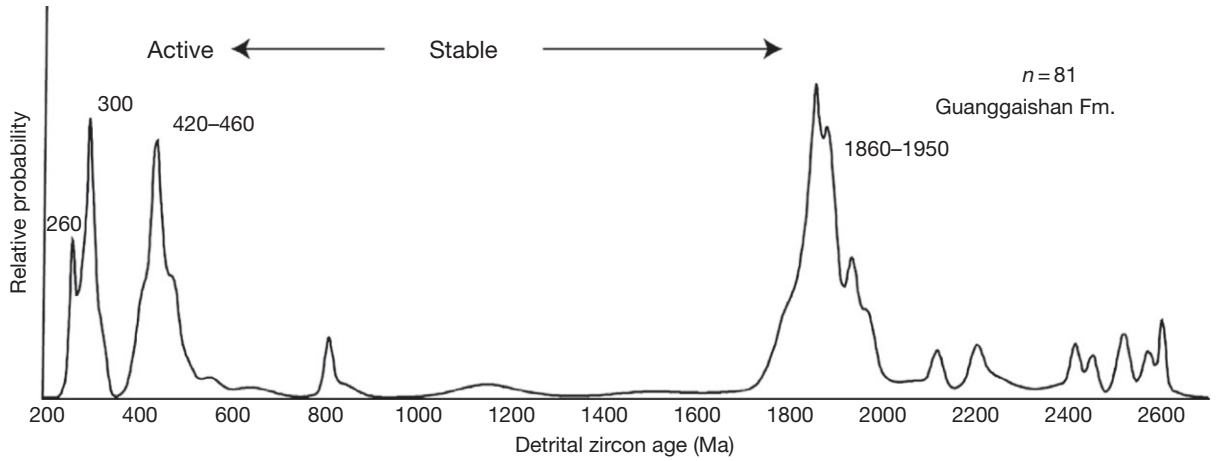


Figure 8 Age distribution of detrital zircons from the Guangganshan formation of Ladianian (Triassic) age in eastern China illustrates problems in obtaining the age of Hadean tectonics from ancient populations. The sample of 81 ages resolves four peaks. The source region was active before the deposition age of ~228 Ma and stable for a billion year interval before then. It is not known whether Hadean source regions represent the relatively stable or the relatively active regions at that time. It is inappropriate to infer global events from this Chinese locality. Archean and younger terrains have had both active and stable episodes. Data from Weislogel et al. (2006).

9.06.3.4 Asteroid Bombardment

Asteroids bombarded the early Earth. The precise timing of the impacts is unclear given the lack of geologic record. Current lunar data are compatible with a relatively quiescent period from 4.4 to 4.0 Ga followed by 'late heavy bombardment' between 3.8 and 4.0 a Ga (see, Bottke et al., 2012; Ryder, 2002; Zahnle and Sleep, 2006; Zahnle et al., 2007). Extrapolating from the Moon constrains the size and total number of objects that hit the Earth. Over 100 of them were greater than 100 km in diameter. About 16 were larger than the largest (170 km diameter; Potter et al., 2012) object that hit the Moon. A few (0–4) may have been big enough (>300 km diameter) to boil the ocean to the point that only thermophile organisms survived in the subsurface. These impacts on Earth brought brief returns of rock vapor and steam greenhouses.

The effect of impacts on tectonics, however, was modest. The Moon, which was subjected to the same bombardment, retains its upper crust over much of its surface. Cavosie et al. (2004) saw no evidence of impacts in their sample of detrital zircons going back to ~4.4 Ga. This includes zircons with metamorphic rims, which might show annealed shock features. Abbott et al. (2012) stated that ~1/3 of the crust present during the bombardment would have been heated enough by impacts for metamorphic overgrowths to form on zircon grains. They found features in their zircons consistent with this process. Conversely, the durations of the initial rock vapor atmosphere and runaway greenhouse were geologically brief and unlikely to preserve rocks. The existence of ~4.4 Ga detrital zircons and ~4.5 continent-like geochemical reservoirs (Harrison et al., 2005) does not bear on the reality of the Moon-forming impact.

9.06.4 Dawn of Plate Tectonics

It is not clear when modern plate tectonics started. We see none of the Earth's original crust, and when we do get zircon samples at ~4.4 Ga, the evidence is compatible with plate-like processes. We see no evidence for any period where the surface of the Earth was a one-plate planet like Mars. The viable candidates are a mush ocean with significant regions of continental crust and lithosphere and plate tectonics with solid lithosphere above solid mantle. As noted already, the transition between these two modes may have well been protracted, regional, and gradual. Various aspects of modern tectonics may have begun at different times.

The widespread occurrence of vertical tectonics in the Archean does not preclude plate processes. Our discussion follows comments by Sleep (2005) and Condie and Benn (2006). As already noted, vertical tectonics occur over broad swaths of the present Earth. Second, the Earth's interior was obviously hotter in the Archean than at present. Volcanic sequences tended to be thicker and the granitic crust was more easily mobilized by mafic and ultramafic intrusions. It is likely that some continental areas, like the Minto province in Canada, formed from remobilized oceanic plateaus (Bédard et al., 2003). Alternatively, the Minto basaltic lithosphere may be a remnant of a mush ocean plate.

Van Kranendonk et al. (2004) and Van Kranendonk (2010) discussed the eastern Pilbara Craton. A prolonged period of crustal formation occurred from 3.47 to 3.24 Ga. Notably, supracrustal rocks on the domes are not extensively metamorphosed. This indicates that the surrounding greenstone belts subsided and that the tops of the domes were never deeply buried. It is incompatible with assemblage as a series of alpine-style thrust sheets. An analogous situation exists in the 2.7-Ga Belingwe Belt in Zimbabwe. Sediments on the flanks of the belt are weakly metamorphosed (Grassineau et al., 2005). This is evidence that the center of the belt subsided in situ without deeply burying its flanks. It is evidence against this greenstone belt being a 6 km thick allochthonous oceanic plateau as proposed by Hofmann and Kusky (2004).

Careful work indicates that horizontal tectonics and vertical tectonics coexisted in the Archean. Hickman (2004) showed that vertical tectonics prevailed in the eastern Pilbara Craton before 2.9 Ga, while horizontal tectonics are evident in the western Pilbara Craton. Lin (2005) demonstrated the formation of a greenstone belt as the keel of a strike-slip deformation zone in the Superior Province. Parmentier et al. (2006) concluded that strike-slip became more dominant with time in the late Archean in the Superior Province.

Modern mantle plumes likely affect both continental and oceanic vertical tectonics. They provide some analogy to the affects of generally hot Archean mantle and the local effects of even hotter Archean plumes. Conversely, geologists studying the Archean are tasked with distinguishing the effects of plumes from those of ambient hot mantle. Herzberg et al. (2010) geochemically associated Archean nonarc basalts from ambient hot mantle with source temperatures similar to those at modern plume-dominated ridge axes that form oceanic plateaus. They associated komatiites with the local affects of plumes. Their thermal models and geochemical determination of source temperatures imply that the role of plumes in the Earth's heat budget and their temperature contrast with ambient mantle have increased since the Archean.

With regard to regional geology, Van Kranendonk (2010) related mainly continental to mainly oceanic plume processes to the Pilbara Craton in Australia. The physics of modern plume interaction with vertical tectonics are too poorly understood to warrant detailed modeling in this review. The Kerguelen Plateau is a mainly oceanic analog (Coffin et al., 2002). There is some continental crust with Elan Bank (Ingle et al., 2002). This feature began ~120 my ago, likely with a starting plume head. The plume track has recrossed the plateau and is now likely beneath it. The Basin and Range province in the United States is essentially continental.

The remainder of this section reviews evidence for the earliest documentation of various plate tectonic features. This task involves both preservation bias and mechanics. Overall, geologists see plate-like features as soon as the Earth has a record that could reasonably provide them.

9.06.4.1 Seafloor Spreading and Oceanic Crust

Almost all oceanic crust eventually subducts into the mantle. The tidbit that survives provides a biased sample. Friend and Nutman (2010) reviewed issues that arise with Archean exposures. Parautochthonous ophiolites from back-arc, forearc, and breakup margins start adjacent to continental crust. These rocks may well show continental chemical affinities (Hofmann and

Kusky, 2004). Thin oceanic crust formed during cool slow breakups is most likely to stay intact and become 'Tethyan' ophiolite. Near-arc crust accretes to become a 'Sierran' ophiolite. These complexes, respectively, may grade into stretched continental crust and arc crust. Conversely, we are not likely to find intact 'normal' Archean oceanic crust if it was in fact as thick as modern oceanic plateau crust (Condie and Benn, 2006; Moores, 2002).

To bring up another uncertainty, mafic rocks formed away from ridge axes mimic ophiolites in tectonically disrupted exposures. The rift zones of major edifices like Hawaii constitute a serious problem. The rifts spread at a few centimeters per year. In a submarine section, pillow basalts overlie dikes and then gabbro (e.g., Okubo et al., 1997). Subaerial pahoehoe may even be difficult to distinguish from submarine pillows at highly deformed and metamorphosed outcrops. Cover tectonics on the flanks of edifices resembles foreland thrusts when viewed locally (e.g., Morgan et al., 2003).

There are certainly abundant pillow basalts and other mafic and ultramafic rocks in the Archean. Sedimentary rocks bearing minerals derived from chromite are also common (Sleep et al., 2012, and references therein). Diamond inclusions derived from subducted weathered chromite even exist (Arai and Ishimaru, 2011). Moores (2002), de Wit (2004), and Friend and Nutman (2010) reviewed candidates for oceanic crust.

One putative ophiolite is the 2.5 Ga Dongwanzi complex in China (Kusky and Li, 2010; Kusky et al., 2001). It is now evident that unrelated Triassic rocks were included in the original proposed complex (Zhao et al., 2007). Kusty and Li (2010) and Kusty and Zhai (2012) have now excluded these rocks. Their remaining exposures have late Archean ages and rocks expected within a dismembered ophiolite.

Blocks of apparent oceanic crust in the Pilbara block are well studied candidates that are older than 3.2 Ga (Hickman, 2004; Van Kranendonk, 2010; Van Kranendonk et al., 2004). Upper crustal rocks at Marble Bar are well preserved, providing an early 3.46 Ga example of carbonate formation by deep marine hydrothermal fluids (Nakamura and Kato, 2004). Kitajima et al. (2001) studied similar nearby 3.46 Ga oceanic basalts at North Pole, Australia. Using boiling relationships, they obtained that the water depth was comparable to modern ridge axes with a best estimate of 1600 m.

The Isua area of Greenland includes rocks that Furnes et al. (2007, 2009) consider to be a 3.8 Ga ophiolite. Friend and Nutman (2010) accepted the interpretation of a dismembered Sierran ophiolite but interpreted these rocks as part of a somewhat younger 3.7 Ga terrain. They associated the putative sheet dikes of Furnes et al. (2007, 2009) with nearby 3.5 Ga dike complex that cuts both the 3.8 and 3.7 terrains.

Abyssal peridotites distinguish conventional oceanic lithosphere from lithosphere formed in a mush ocean. This rock type occurs as soon as there is geologic record at 3.8 Ga in Greenland (Friend et al., 2002). Geochemical studies indicate that the serpentinite at Isua formed in a forearc environment similar to the modern Marianas (Pons et al., 2011), providing another analogy to plate processes.

9.06.4.2 Transform Faults

The strike-slip motion on transform faults obviously dissipates energy but provides no gravitational driving force (Sleep, 1992). Mathematically, strike-slip is toroidal motion around a vertical axis. Such motion cannot occur unless rheology is strongly laterally heterogeneous, that is, rigid plates and weak plate boundaries. Polat and Kerrich (2006) summarized known Archean strike-slip faults. As already noted, strike-slip motion is not easily recognized in the presence of associated vertical tectonics (Chen et al., 2001; Lin, 2005; Parmentier et al., 2006). Conversely, local rheological contrasts between domains of basically vertical tectonics may impose a component of horizontal shear on dip-slip faults (Erickson, 2010).

Continental margin transform faults preserve a record of oceanic tectonics on their outboard sides (Sleep, 1992). Their length provides a minimum limit of plate size. The duration of sense of slip constrains the ratio of plate size and plate velocity. One could get the relative plate velocity directly with piercing points or with the demise of volcanism as a Mendocino-type triple junction passed. The gross endeavor may be feasible. Atwater (1970) applied these tectonic concepts to explain Tertiary California geology. There are limits to resolution. Coastal California is quite complicated when examined in detail (Wilson et al., 2005).

The Inyoka fault in Southern Africa is the oldest 3.2 Ga well-preserved example. It cuts syntectonic basins of the Moodies group (de Ronde and de Wit, 1994; Huebeck and Lowe, 1994). Like the San Andreas Fault, motion is parallel to the gross strike; piercing points are not recognized in spite of detailed mapping (de Ronde and de Wit, 1994; de Ronde and Kamo, 2000). Motion is associated with detachment faulting (Huebeck and Lowe, 1994; Kisters et al., 2003).

9.06.4.3 Subduction and Continental Lithosphere

We have already noted that igneous rocks analogous to arc magmas are ubiquitous back to the time ~4.5 Ga of the first hint of a geologic record (Harrison et al., 2005; Polat and Kerrich, 2006; van Hunen and Moyen, 2012). Foundered crust within a mush ocean is likely to be analogous to the crust in modern slabs. Thrusting and rifting are likely to resemble the modern analog at shallow depths of exhumation as brittle failure is involved. On the other hand, Stern (2005) contended that the lack of high-pressure and ultrahigh-pressure metamorphism on the ancient Earth indicates that modern subduction did not occur. Rocks from these processes in fact are more common in younger terrains but date back to at least ~3.0 Ga (Brown, 2006).

Subduction, moreover, merely requires that surface rocks penetrate to great depth, while ultrahigh metamorphism requires that these rocks return to the surface in a recognizable form. The higher temperature in the Archean mantle clearly heated subducted material. Any subducted material that made it back to the surface was likely to be hotter than modern returning material (Condie and Benn, 2006).

We journey downward into deep cratonal lithosphere in search of a better rock record. Lithospheric mantle beneath cratons is less dense than normal mantle. This along with its higher viscosity keeps continental lithosphere from being entrained into the deeper mantle by stagnant-lid convection (Doin et al., 1997; Sleep, 2003). Its great yield strength guards it during continental collisions (Lenardic et al., 2003). The oldest in-place diamond pipe is 2.7 Ga in the southern Superior Province of Canada (O'Neill and Wyman, 2006). Detrital

diamonds of that sedimentary age and likely eruption age exist nearby (Kopylova et al., 2011). Detrital diamonds in the 2.5-Ga Witwatersrand sequence also indicate that thick stable Archean lithosphere had already cooled into the diamond stability field by that time (Nisbet, 1991, p. 61).

For example, kimberlite eruptions exhumed late Archean eclogite xenoliths in Greenland (Tappe et al., 2011). Eclogitic facies diamonds globally are less than 3.2 Ga and become dominant by 3.0 Ga, raising the possibly of a transition toward modern plate tectonics (Shirey and Richardson, 2011).

The Archean subsurface crystallization age of xenoliths in diamond pipes and diamond inclusions is well known. For present purposes, sulfides in ~2.9 Ga diamond inclusions from South Africa studied by Farquhar et al. (2002) are relevant. The sulfides have mass-independent sulfur fractionation. This indicates that this sulfur was exposed to UV radiation in an anoxic (<2 ppm O_2) atmosphere (Pavlov and Kasting, 2002). The sediments and altered volcanic rocks containing this sulfur were subducted into the ~200 km deep lithosphere. This material remained sequestered until tapped by a kimberlite.

In addition, some deep lithospheric samples provide evidence that their depths of formation were shallower than their depths of sequestration (Canil, 2004; Jacob et al., 2003; Menzies et al., 2003; Saltzer et al., 2001; Schulze et al., 2003a,b). This feature and the chemistry of the samples indicate entrainment to depth in the wedge above the slab. Conversely, they do not require highly elevated temperatures that would generate komatiites and their residuum at ~200 km depth. Herzberg and Rudnick (2012) concentrated on mantle samples that had not been altered by later processes. They concluded that cratonal lithosphere formed in a nonarc environment at elevated mantle temperatures analogous to modern oceanic plateaus.

Overall, a process with the net effect of subduction was required to sequester most of its CO_2 into the mantle and render the surface habitable before ~3.9 Ga. CO_2 and water entered the mantle with altered oceanic crust and sediments. The concentration of these materials in subducted domains is then >100 times the bulk mantle. This material cycled through the mantle and remelts forming CO_2-rich rocks including kimberlites. The oldest such example from India had a complicated history (Upadhyay et al., 2009). Measurement of $^{142}Nd/^{144}Nd$ indicates the source region of these rocks became isolated, likely from seafloor spreading, seafloor alternation, and subduction at ~4.2 Ga (adjusted to ~4.3 Ga for the revised half-life of ^{146}Sm in the work of Kinoshita et al., 2012). However, Roth et al. (2013) were unable to replicate these ^{142}Nd measurements. We continue discussion to illustrate the implication of any eventual similar valid measurements. At ~3.6 Ga, igneous rocks derived from partial melting of this material intruded into the deep cratonal lithosphere of India. At 1.48 Ga, the deep lithosphere remelted producing igneous intrusions at shallow depth. Overall, the mantle may sequester a record of last resort of Hadean subducted material (Sleep et al., 2012).

9.06.4.4 Subduction and Continental Crust

The formation of continental crust was most vigorous in the Archean between 2.5 and 3.5 Ga (Herzberg and Rudnick, 2012). Mass balance constrains mechanisms for the formation of this crust. Continental crust is obviously not composed of high MgO magmas that are complementary to cratonal lithospheric mantle. Such magmas are instead minor constituents of continents. Mafic material complementary to the formation of continental crust from mantle-derived melts has foundered back into the mantle (Herzberg and Rudnick, 2012). Early continental crust did not fully blanket the Earth, as mantle-derived igneous rocks without continental signatures are common. The continental crust contains ~1/3 of the bulk silicate Earth's inventory of, for example, K, Th, and U, so a large fraction of the mantle must have been involved in the formation of continental crust over time.

Figure 7 illustrates the net effect (but not the unknown precise geometry) of continent-forming processes envisioned by petrologists that involve regions of hot thickened mafic crust. Basaltic crust thrust into the hot gabbro melts and the melt ascends without encountering the mantle wedge that exists above normal slabs at modern arcs. Modern arc magmas react with the mantle wedge or are derived directly from partial melting within the wedge, acquiring nickel and chromium in both cases.

Igneous rocks derived from thickened modern Jamaican oceanic plateau provide Archean analog. They contain much less Ni and Cr than typical arc magmas as expected from the local absence of a mantle wedge (Hastie et al., 2010a,b). In general, the common class of Archean granitic rocks (trondhjemite, tonalite, and granodiorite, TTG suite) contains little Ni and Cr (e.g., Hoffmann et al., 2011). Geochemical studies indicate that the sources of the TTG suite were nonarc plateau-like crust (Condie, 2008; Herzberg and Rudnick, 2012), long-lived early Hadean mafic crust (Næraa et al., 2012), and arc tholeiities (Hoffmann et al., 2011; Nagel et al., 2012).

Geochemistry bears on the depth of the source region of TTG suites. For example, garnet remains in the source region at higher pressure and retains heavy rare earth elements but not light ones. The resulting melts are enriched in light rare earths but depleted in heavy ones. Both moderate depth source regions (30–45 km) associated with the lower crust and deeper source regions compatible with subduction into mafic material are indicated (Hoffmann et al., 2011; Moyen, 2009; Nagel et al., 2012).

The process envisioned earlier is basically that described for nonplate Archean tectonics within thickened plateau crust (Bédard et al., 2012). The main difference is that regional TTG-forming processes may have coexisted with plate-like processes. Plumes are attractive mechanisms to form plateau-like crust in the first place and to later mobilize thickened deep mafic crust. Dynamic modeling of Archean thickened mafic crust has recently begun (Sizova et al., 2010; van Hunen and Moyen, 2012). It is likely that thickened mafic crust made subduction episodic (van Hunen and Moyen, 2012). Section 9.06.5 discusses constraints on the rate of plate tectonics.

With regard to regionality, it is geometrically unlikely on the spherical Earth that hot plateau-like crust arrived everywhere alone a subduction at the same time. Cold old slabs drove hot plateau convergence elsewhere along the plate boundary. There is some evidence that large plates existed on the early Archean. Boninites erupted along the modern Izu–Bonin–Mariana forearc, where the old crust of the large Pacific

Plate subducts. Turner et al. (2014) documented a suite of similar >3.8 Ga rocks in Quebec.

Episodicity, regionality, and the coexistence of different tectonic processes make it difficult to estimate globally averaged heat flow and even to state that plate tectonics was the dominant heat flow mechanism. Furthermore, regionality implies boundaries existed between active regions of thickened crust and inactive continental and oceanic crust. Mechanically, hot thick mafic lower crust was buoyant and would have spread laterally unless 'backstops' were present. Hot buoyant mantle plume material would also spread laterally. Slabs, especially on their arc sides, are obvious backstops for both crustal mafic and mantle plume materials. Locally thin regions of mantle lithosphere particularly where it was already cool, depleted, and chemically buoyant would have acted as catchments for both types of buoyant material. Cool continental crust is a reasonable catchment for more fluid hot crust. We do not have a modern analog of hot mafic lower crust impounded within a continental area.

9.06.5 The Rate of Plate Tectonics over Time

The record after 3 Ga becomes increasingly interpretable in terms of the rates of plate motions. One would like to deduce the change in the temperature of the Earth's interior over time along with the variation in the evolution of plate motions. In the absence of magnetic lineations, methods grab at straws.

9.06.5.1 Longevity of Geologic Features

A direct approach utilizes the longevity of geologic features (Bradley, 2011). We already did this in **Section 9.06.3.3** with the clusters and gaps in detrital zircon population ages (**Figure 8**). In particular, Thern and Nelson (2012) deduced gaps of 160 and 300 my between zircon-forming activity and docking to two terrains in Australia at ~3.747 Ga.

Application of this method becomes simpler after 3 Ga in that one can more easily associate a tectonic event with a radiometric age. The length of the Wilson cycle between passive margin breakup, to arc collision, and final continental collision is attractive. The interval between supercontinent assemblies is a more macroscopic aspect. We have already mentioned the duration of strike-slip events on active margins. The formation to accretion duration of seamounts and the ridge to trench lifetime of ophiolites provide direct information on oceanic plates (Moores, 2002).

Korenaga (2006) and Herzberg et al. (2010) briefly reviewed this evidence. Additional data since Sleep (1992) discussed tectonic rates provide better constraints (Bradley, 2011). The main difficulty is that plate velocities and the durations of margins, ocean crust, and seamounts are highly variable on the modern Earth. Other than that, rates have not systematically varied by a factor of 2 from current rates over the last 3 Ga; the author does not attempt to decipher a trend. Korenaga (2006) and Herzberg et al. (2010) contended that rates have actually increased since 3 Ga. Bradley (2011) favored that Proterozoic 545–2500 Ga rates were faster than current and Archean rates.

Paleomagnetic data provide constraints on mainly continental plate velocities. For example, Blake et al. (2004) obtained precise radiometric ages on strata in the eastern Pilbara Craton for which paleolatitudes had been determined. Samples at 2721 ± 4 Ma and 2718 ± 3 Ma differ by 27° or 3000 km. Taken at face value, the drift rate was 1000 mm $year^{-1}$. The drift rate was at least 100 mm $year^{-1}$ within the error of the data. The faster rate is ~10 times modern rates. Suganuma et al. (2006) studied the Archean Marble Bar Chert also from the Pilbara Craton. Over a ~20 my interval around 3.46 Ga, the samples apparently moved 21°. Within nominal uncertainty, the drift rate was 112–1120 mm $year^{-1}$. Sleep and Windley (1982) hypothesized brief episodes of rapid Archean plate motion when old thick lithosphere subducted into the underlying fluid mantle. van Hunen and Moyen (2012) favored episodic subduction rates. Available plate size (here at least 3000 and 2400 km) ended the rapid drift episodes often by continental collisions. Transient mush oceans would have a similar effect, but local regions of hot thick mafic crust might have the opposite effect. Such periods of rapid drift (if real) need to be included in the global average plate motion.

9.06.5.2 Degassing and Heat Balance

The partition of rare gases between the atmosphere and the mantle constrains the rate of tectonic processes over geologic times. The traditional approach uses isotopes of argon (Padhi et al., 2012; Sleep, 1979; Tajika and Matsui, 1993). Decay of ^{40}K produces ^{40}Ar, with a half-life of ~1.25 Ga. This isotope is a significant component of the atmosphere but is also present in mantle-derived gases. The other isotopes ^{36}Ar and ^{38}Ar are primordial. They are strongly concentrated in the atmosphere compared with the mantle. That is, these isotopes degassed before ^{40}Ar formed, most likely in the aftermath of the Moon-forming impact. They have not been subducted in significant quantities since that time. The mantle conditions in the aftermath of the Moon-forming impact have never returned. Early argon escape does not provide a strong constraint on the behavior of argon once the mantle had cooled and plate tectonics had commenced.

The heat balance of the Earth in **Section 9.06.2.1** combined with degassing of the Earth from partial melting at ridge axes in **Section 9.06.2.5** constrains both the thermal history and the tectonic rate of the Earth. One assumes that argon is an incompatible element that nearly quantitatively enters magma and degasses at ridge axes (Cassata et al., 2011; cf. Watson et al., 2007). The flux of argon into the atmosphere is proportional to the thickness of mantle that melts at ridge axes, hence to the temperature of the mantle and to the global rate of seafloor spreading. Qualitatively, a significant fraction of the Earth's ^{40}Ar remains in the mantle; convection in the past has been sluggish since it was generated.

This reasoning is quantified since the ratio of heat-producing elements in the mantle K:Th:U is reasonably constrained from sampling of mantle-derived lavas, assuming that no 'hidden' reservoirs of argon or heat-producing elements exist in the mantle. The total production of ^{40}Ar is thus proportional to the Urey number. A high value of the Urey number implies that a large amount of ^{40}Ar has been produced throughout geologic time. It also implies vigorous convection

in that this radioactivity did not heat the mantle to very high temperatures. The predicted amount of degassed ^{40}Ar greatly exceeds that in the atmosphere. A modest Urey ratio as in Figure 3 implies a reasonable amount of ^{40}Ar production and a reasonable partition of it between the mantle and the atmosphere. The rate of plate tectonics is comparable to the present.

Padhi et al. (2012) examined the xenon system. Decay of ^{129}I produced ^{129}Xe with a half-life of 15.7 my; the minor fission branch of ^{244}Pu decay produced ^{136}Xe. Its half-life is 82 Ma, mainly from alpha decay. The minor fission branch of the long-lived isotope ^{238}U also produced some ^{136}Xe as well as other xenon isotopes; analyses of the suite of Xe isotopes separate its contribution from ^{144}Pu. Most of the global inventory of xenon is in the atmosphere indicating vigorous early degassing (Yokochi and Marty, 2005). However, the mantle has excess iodine- and plutonium-derived xenon compared with the atmosphere. Hence, tectonics became sluggish within few half-lives of these parent isotopes after the Moon-forming impact. The long-term argon arguments carry through since the initial ratio Pu:U is known so that total heat production is proportional to initial Pu concentration. That is, the Urey number is modest and the rate of plate tectonics has been comparable to the recent one.

9.06.5.3 Freeboard and Stress Within Plates

The freeboard above sea level of the continents provides both a kinematic constraint and a dynamic constraint. Basically, many Archean terrains have never been deeply eroded or deeply buried by sediments once they were stabilized (e.g., Galer and Mezger, 1998). The elevation of these regions relative to sea level thus has not changed much. If the amount of seawater has been relatively constant, the available room in ocean basins determines sea level. Thick Archean crust is more buoyant than modern crust and should cause a sea-level rise. Sea level rises as global continental area increases. Vigorous plate tectonics implies that the oceanic basin is filled with young shallow crust. This causes a sea-level rise. Overall, the data are compatible with plate rates comparable to present since 3 Ga.

Mechanics provide a more reliable version of this constraint that does not involve past ocean volume (England and Bickle, 1984). The initial thickness of Archean continental crust is similar to that of young crust. Continental crust is buoyant and would spread laterally like oil over water if it were fluid. A compressional orogeny must overcome this force to form the continent. The available force scales to the ridge push force in eqn [5] and hence to the age of seafloor at subduction t_E. The lack of change in compressional orogenies since 3 Ga indicates that rate of plate overturn has not changed much.

9.06.6 Death of Plate Tectonics

I have already shown that the rock vapor atmosphere and the mush ocean ceased when the available energy flux failed to maintain them. Plate tectonics will meet the same fate as the mantle cools. The process, called ridge-lock, involves melt generation at ridge axes. If the interior temperature is lowered by 50 K from its present value, much less melt is generated (see eqn [25]). Solid peridotite rather than a fluid mush chamber and sheet dikes must deform for spreading to occur.

In the model (Figures 1–3), the mantle becomes too cool for plate tectonics in the future at -0.9 Ga. The heat flow drops to the stagnant-lid branch (Figure 1, point E to point F). The mantle slowly heats up but does not become hot enough to restart plates. At −2 Ga, radioactivity heating is in equilibrium with heat flow. The temperature peaks (Figure 3). The mantle cools thereafter without returning to plate tectonics. Alternatively, the transition is gradual. The mantle stays at the transition temperature until radioactive heat generation and heat flow balance, again at −2 Ga (Figure 3).

Two harbingers of doom for plate tectonics exist on the present Earth. First, full spreading rates are less than 20 mm year^{-1} along the Arctic spreading center. This rate is slow enough that material cools as it ascends, violating the adiabatic ascent assumption in deriving eqn [25]. Much of the crust is serpentinite formed from hydrated peridotite (Dick et al., 2003). Gabbros freeze at depth. Convection then carries them to the surface. The radiometric ages of such rocks can be millions of years older than the nominal age of the seafloor.

Second, the Australian–Antarctic spreading center produces a 'discordance zone' with thin ocean crust and rugged topography. The crustal thickness and the unusually large depth at the ridge axis indicate that the upwelling mantle is cooler than that at normal ridges. Ritzwoller et al. (2003) presented tomography showing a reduced potential temperature of the upwelling mantle of ~100 K relative to normal ridge axes. They compile petrologic reduced temperatures of 60–150 K. Herzberg et al. (2007) reviewed the global variation of magma source temperatures.

Both the Arctic ridge and the Australian–Antarctic discordance are modest segments of boundaries between major plates. Forces from the rest of these plates could well supply the additional driving force for spreading. When the Earth's interior has cooled, both slow spreading and cool upwelling will tend to lock the entire length of all ridge boundaries. Neither Mars nor Venus provides even a qualitative analogy at the present level of ignorance.

A second more subtle effect is already occurring (Sleep, 2005). Stagnant-lid convection provides heat flow to platform regions where the deep lithosphere is not chemically buoyant. The lithospheric thickness in these regions from eqn [13] has increased as the mantle cooled. It is now similar to that beneath cratons where chemical buoyancy defines the base of the lithosphere. Further cooling of the mantle will bring the rheological boundary layer in eqn [14] below the zone of chemical buoyancy everywhere on the planet. The stagnant-lid versus chemical-lid distinction between platforms and cratons then will cease to exist.

9.06.7 Biological Implications

The minimal requirements for the survival of life, energy sources, and habitable temperatures are linked. Non-photosynthetic life couples a Gibbs energy-producing reaction, like methanogenesis, with an energy-consuming reaction, like ADP ⇒ ATP. High temperatures speed reaction abiotic rates, making energy-producing kinetically inhibited reactions less

available. They tend to cause the decomposition of the organic matter within all life-forms, both squandering the energy that the organism has gone to great ends to convert into a useful form and trashing its genetic information. The geologic history and processes discussed earlier are relevant to the establishment of biologically favorable conditions and affect of biology on geologic processes and the future of life on the planet.

9.06.7.1 Origin of Life and Its Survival on the Early Earth

I follow the discussions of Sleep et al. (2001), Zahnle et al. (2007), Sleep (2010), Sleep et al. (2011), and Arndt and Nisbet (2012) to put the establishment of habitable conditions into a tectonic context. The Earth cooled to a solar greenhouse within several million years after its formation. At first, over 100 bars of CO_2 maintained temperatures above 200 °C. Carbonates were stable within a thin rind at the surface. The volume of the rind, however, was too small to take up a sufficient quantity. At some stage (4.44 Ga in the model), global heat flow waned to 1 W m^{-2}. The Earth resembled oceanic crust younger than 1 my, the age t_E in eqn [2] of subduction at that time. The crust was a significant CO_2 sink. Carbonate formation now occurs within the uppermost several 100 m of young oceanic crust. Mass balance indicates that crustal overturn was necessary to sequester the CO_2. Available divalent cations in the uppermost crust could take up only ~1/6 of the global CO_2 inventory. Degassing of foundered crust returned an unknown amount of CO_2 to the surface. With a global resurfacing time of 1 my, 6 my was needed to completely consume the dense CO_2 atmosphere in the absence of any degassing or return flux. Alternatively, plate rates were sluggish once CO_2 could deeply subduct. A clement ~1 bar CO_2 atmosphere could have existed over a pH ~6 for a geologically and biologically significant time (Sleep et al., 2011).

Once the primordial CO_2 was sequestered in the subsurface, ejecta from impacts were an effective CO_2 sink. The climate was freezing until life produced methane, a potent greenhouse gas (Sleep and Zahnle, 2001). Life could well have arrived within Mars rocks. Mars was habitable up to 100 my before the Earth. Martian origin is more testable than Earth origin. The Mars record is intact, just hard to retrieve.

After the demise of the thick CO_2 atmosphere, the Earth was habitable except for brief (10^4 years) periods after asteroid impacts (Sleep and Zahnle, 1998; Zahnle and Sleep, 2006; Zahnle et al., 2007). A few impacts large enough to vaporize much of the ocean are likely, but not certain. They left only thermophile survivors below 1 km deep in the subsurface. Terrestrial life could well root in these survivors. Molecular genetics is compatible with a sequence where life originated at clement temperatures, colonized the planet including the deep subsurface, and survived a thermophile bottleneck (e.g., Boussau and Gouy, 2012).

The main safety requirement in the aftermath of impacts is that multiple areas with thermal gradients less than 100 K km^{-1} existed on the planet. For a hard-rock conductivity of 2.4 W m^{-1} K^{-1}, this requires a heat flow less than 0.24 W m^{-2}. This is the heat flow within older oceanic crust when the average heat flow is 0.48 W m^{-2}, occurring at 4.37 Ga in the model. Continent-like regions, however, date from ~4.5 Ga. These low heat flow regions could well have provided earlier refugia.

Bottke et al. (2012) proposed the Earth experienced a minimum in impact flux following the tail of accretion. Later, ~4.1 Ga orbital perturbations in the outer solar sent asteroids into Earth-crossing orbits. If this hypothesis is correct, there was time for cool subsurface refugia to develop before dangerous impacts occurred. In agreement, observations of Archean impact spherule beds indicate that basin-forming impacts of ~50 km projectiles continued well into the Archean (Johnson and Melosh, 2012).

Life by 3.85 Ga had dispersed globally. The descendants of the organisms that survived asteroid impacts placed less onerous conditions on tectonics over the remainder of geologic time. They mainly demanded that tectonics recycled volatiles and nutrients so that the planet did not end up like Venus or Mars.

9.06.7.2 Biological Effects on Geologic Processes

The effects of life on continental and even mantle processes are significant (Sleep et al., 2012). It is obvious that life affects chemical weathering and leads to the standard classes of sediment, sandstones, shales, and chemical deposits including carbonates. Metamorphism transforms these deposits into granitic continents (Rosing et al., 2006). Furthermore, biology affects the composition of the uppermost oceanic crust and marine sediment. The subduction megathrust slips along these materials that later dehydrate and remelt to form arc volcanics. Some of the subducted material continues into the deep mantle and later remelts when brought back near the surface. Over geologic time, these processes have partitioned C, N, and S between the crust and mantle.

Biological sediments types have varied throughout geologic time. Over the last ~150 my, the shells of pelagic organisms have accumulated on shallow seafloor near ridge axes (e.g., Sibley and Vogel, 1976). Little of this material has yet to reach subduction zones (Plank and Langmuir, 1998). It is thus not obvious whether it will accrete or be deeply subducted. In this regard, Saura et al. (2008) described marine carbonates accreted onto Cuba.

Banded iron formation became rare after ~1.85 Ga when the deep ocean became oxic (e.g., Slack and Cannon, 2009). However, no chemical domains in the mantle and varieties of arc volcanics have been recognized as forming at its expense. Dobson and Brodholt (2005) suggested that banded iron formation became dense mantle regions just above the core–mantle boundary. Retention of some banded iron formation in this dregs region is reasonable, but the strong form of this hypothesis does not pass mass balance constraints. The mass of deep-mantle dreg regions is much greater than that of the continent crust and hence any chemical sediments formed at its expense.

Overall, life based on anoxygenic photosynthesis teemed at ~3.85 Ga, the time of the first preserved remnant of sedimentary rocks. Intact rock as opposed to zircon crystals older than this time is extremely rare. The question arises whether this synchrony is merely a coincidence. It is conceivable that photosynthesis greatly predates the rock record. It is also conceivable that biological processes associated with

photosynthesis-based ecosystems modulated tectonic and geochemical processes facilitating rock preservation.

Geologists should vet putative prebiotic environments to see whether they could actually have predated life. Conversely, firm biological conclusions would show that certain situations actually existed on the early Earth. As a brief example with explicit geology, Mulkidjanian et al. (2012) proposed that life originated on 'dry' land in an arc-like environment. The oldest Archean record contains arc-like rocks and their weathering products. Hadean zircons indicate rocks crystalline formed at the expense of land-derived sediments. Biologically attractive aspects of this setting include that tectonics interleaved a varied of rock types and that hydrothermal circulation penetrated these rocks. Evaporation and wetting along with freeze and thaw alternatively concentrated and diluted organic solutes. Moreover, the bases present in DNA and DNA, cytosine (C), thymine (T), uracil (U), adenine (A), and guanine (G), are relatively stable compared with plausible alternatives in the presence of UV light (Serrano-Andrés and Merchán, 2009). This trait has obvious selective advantage on land where UV is present in sunlight, not at deep marine vents (Mulkidjanian et al., 2012).

9.06.7.3 The Extraterrestrial Export of Biological Concepts and the Future of Life

Terrestrial organic evolution is a mixed blessing for inferring what happened on other planets. One cannot blindly extrapolate cosmically from the period where we have a good geologic record, especially since life affects geologic processes. Life has evolved to be fit for the conditions it actually experienced during the Earth's history. For example, the Earth has the right amount of water so that tectonics and erosion both tend to generate surfaces near sea level (Sleep, 2005). Abundant shallow seas made the transition to land easy. Overall, evolution for fitness to current conditions produces the illusions of design and providence. Any intelligent organism, as a survivor, will find that its personal family tree and its species have had a harrowing history, giving the illusion of good fortune to the point of miracle.

Past performance does not imply future results. Tectonic vigor wanes. Our Sun waxes. Unless a high-tech society intervenes, our greenhouse will runaway because water vapor is a potent greenhouse gas (e.g., Kasting and Catling, 2003). An increase in global sea surface temperature produces more water vapor and hence a hotter climate. Above a threshold, the feedback is unstable. Our life-giving ocean will be our curse.

Tectonics may extricate us from disaster by jettisoning the ocean in the mantle. The mass balance is reasonable. Right now, basalt covers about 90% of the ocean's surface and serpentinite covers the other 10%. Basalt hydrates at ridge axes to form greenstone. This is a modest water sink. An equivalent thickness of 1–2 km is hydrated to 3% water by mass or 10% by volume (Rüpke et al., 2004; Staudigel, 2003). The basalt flux by slabs is equivalent to 100–200 m of ocean depth globally every 170 my. Serpentinite is ~40% water by volume (Rüpke et al., 2004). Its 10% of the seafloor carries an equivalent of 40–80 m every 170 my. In total, it takes 1.5–3.0 billion years to subduct the ocean equivalent thickness (2500 m) of water. However, most of this water returns to the surface in arc volcanoes (Wallace, 2005).

A prolonged period of sluggish seafloor spreading occurs in the models (**Figures 1–3**). The Earth may transition to rather than jump to the stagnant-lid mode of convection. The cool mantle during this epoch implies that much of the seafloor will be serpentinite and that any basaltic crust is highly fractured. The mantle will be too cool for most arcs to be an effective water-return mechanism. The model spreading rate at −0.9 Ga in the model is still 83% of the current rate at the transition. A fully serpentinized ocean would subduct the available water in 0.65–1.3 Ga at that rate.

The Earth would resemble the fictional planet *Dune* once the oceans have been subducted (Abe et al., 2011). There will still be some ground water. There will be open water and ice only at the poles. This cold trap will set the global dew point to a low temperature. The meager amount of water vapor will be an ineffective greenhouse gas. Our descendants may well help these processes by pumping water into serpentinite and harvesting H_2 gas from the effluent. The Saturn moon Titan is in an analogous state with its greenhouse gases, ethane and methane, trapped at the poles.

On the other hand, plate tectonics may end suddenly and soon. In that case, erosion will bevel the continents. Some volcanoes and dryland may persist, but not the bounty associated with active tectonics. We will not need a set to remake *Waterworld*.

9.06.8 Conclusions and Musings

Geologists' view of the ancient Earth and their repertoire of methods have changed beyond recognition since 1963, when the author took his first college-level geology class. Geochemistry and geochronology were mostly aids to mapping rather than powerful quantitative techniques directed at physical processes. The idea that high-temperature metamorphism was exclusively Precambrian flourished. There was a general hesitancy for considering geology beyond a local scale. Physics did not seem relevant. Geosynclines produced orogenies whenever and wherever they pleased.

Geodynamic models of the early Earth were generic in the 1970s. They did not utilize available information including the presence of high-temperature komatiite lavas in the Archean. 'Don't think' signs remained posted around the Archean 'basement complex' until the late 1970s. For example, a senior faculty member criticized the author for pointing out transposed dikes on a departmental field trip to the Superior Province. These features, when of venerable Archean age, were to be called by the uninformative generic term 'schieren.' Brian Windley visited Northwestern right after this trip. He had much better, more deformed, and far older examples in Greenland that he traced into intact dikes. We began to think about the Archean in a quantitative physical manner.

The high-hanging fruits of years of careful fieldwork with geochronology and geochemistry are now in the basket. There is meager preservation of the oceanic record. Still, geologists often recognize the basic features of plate processes when they look in the Archean. However, Archean tectonics did differ from the modern process. The deep continental crust

frequently acted as a fluid. Dense mafic rocks foundered into the lower crust. Broad areas like the Minto block in Canada and the eastern Pilbara Craton were dominated by vertical tectonics.

There is enough information to rationally discuss the evolution of tectonic rates. Tiny zircon crystals show that continent-like material formed and survived soon after the Moon-forming impact. Continent-like areas continually remelted. Episodic tectonics occurred at given continental localities from the Archean to the present.

At ~3 Ga, the convecting mantle reached a maximum temperature ~200 K hotter than at present (**Figure 3**; Herzberg and Rudnick, 2012). The mantle likely overcooled in the aftermath of the Moon-forming impact to ~100 K greater than its current temperature. Sluggish plate-like processes did not fully vent the heat supplied by radioactivity. Continent formation was most vigorous in the 2.5–3.5 Ga period about the thermal maximum. The possibility arises that later Hadean continental formation was sluggish as the mantle was relatively cool. Kemp et al. (2010) and Arndt and Nisbet (2012) envisioned continual turnover of granite continental material. However, this process was driven by radioactivity within the granitic material not heat from the mantle. Hadean zircon populations do indicate local periods of little zircon production (Thern and Nelson, 2012).

There is enough rock record after 3.0 Ga to meaningfully talk about the rate of plate processes. The answer may not be satisfying. The global rate of plate tectonics was comparable to the present rate. It remained so thereafter.

Chemically buoyant continental lithospheric mantle kept some overlying continental crust out of harm's way. Its basal region sequestered a record of the conditions at the time of its formation that ascends to the surface as xenoliths. Geochemists are now examining this treasure trove. Their work establishes the antiquity of subduction. At a minimum, subduction of basaltic crust into the mantle is required to sequester the Earth's CO_2.

Crameri et al. (2012) showed that familiar one-sided subduction is a consequence of the gradual bending of the oceanic plate into the slab. Two-sided subduction in numerical models is an artifact of inappropriate boundary conditions. The geologist can expect the feature of one-sided subduction including island arcs and accretionary complexes in the deep past. Still, current dynamic models do not even do a good job of representing the spontaneous formation of modern plate boundaries. A damage rheology where faults form and heal is a minimum requirement (Tackley, 2000a,b). In this case, the instantaneous rheology implies an increase in stress for an increase in strain rate. The steady-state rheology is strain rate strengthening below a critical strain rate and 'yield' stress. It is strain rate weakening above that stress and strain rate. This represents the observation that plate boundaries are weaker than typical rocks failing in friction (Zoback and Townend, 2001). The numerical models will need to nicely represent 'normal' plate boundaries, hot-spot ridges like Iceland, mobile continental crust, arc-continent collisions, oceanic subplate boundaries, and oceanic plateaus before they can be exported to the ancient Earth and to other planets.

The simple quantitative models in this chapter provide heat balance constraints and a guide to possibilities. Lenardic and Crowley (2012) outlined the limitations of geodynamic theory. Multiple stable modes of convection may be possible for a given planet and contingencies, like the Moon-forming impact for the Earth, and may determine the sequence of modes. The feedback between global heat flow and continental area has only recently been studied numerically (Lenardic et al., 2005). The water content of the mantle may have a significant effect (Korenaga, 2011).

Conversely, an apparently physically consistent and reasonable model may well yield on inapplicable thermal and tectonic history. Specifically, the three models in **Figure 3** all imply acceptable thermal histories within the scatter of the data of Herzberg et al. (2010) and Herzberg and Rudnick (2012). The nonmonotonic model and the Herzberg et al. (2010) both have a thermal minimum after the Moon-forming impact (**Figure 3**). The models differ in their tectonic rates; the nonmonotonic model has multiple subsequent peaks in tectonic rates and mantle temperature, rather than a broad temperature peak around ~3.0 Ga with relatively constant plate rates. There are hints of such multiple peaks in crustal production in the geologic record (Bradley, 2011), but their significance is unclear. The author concludes that global transitions back into mush ocean are unlikely, but conceivable.

Thus, the geologic record is the master and geodynamics is the guide. The transition from mush ocean to plate tectonics was complicated. Before the transition, partly molten material at the base of the mush ocean facilitated convection in the underlying mantle and foundering of cooled mafic crust. After the transition, the need for the base of the oceanic crust to freeze hindered subduction.

Now that we have geologic record, Archean tectonics are more complicated than modern plates. Continental tectonics often involved fluidlike deformation in the crust initially at the expense of thick oceanic plateaus. The Earth returned regionally, but not globally, to mush ocean conditions, during the Archean thermal peak. Basaltic crust subducted into hot gabbroic mush. The TTG suite of granitic rocks thus produced dominates Archean exposures. Its low Ni and Cr contents indicate that these melts did not pass through or form within the mantle wedge. As at present, slabs subducted into the mantle above arcs, but left much less geologic record of their existence.

Analogous situations of the modern Earth are rare, like feeble volcanism in Jamaica (Hastie et al., 2010a,b). Broad vertical tectonic regions like the Basin and Range do exist within continents, here from the impingement of hot plume material into thin back-arc lithosphere. However, continental mafic rock (here arc volcanics, the Columbia basalts, and Snake River Plain) has not subducted into the lower crust. Collision of a continent or large submarine plateau in the last ~16 my with the Cascadia margin would have caused inboard compression and vertical tectonics. Thick Columbia and arc basalts could well have foundered aided by negative buoyancy in analogy to Archean events. This hypothetical outboard evidence might not be evident billions of years in the future.

Generalizing from the observation that plate-like behavior occurred on the Earth for much of its history, plate tectonics is a viable candidate mechanism on other worlds. Pressure-release melting clearly occurs over an extensive depth range

in the low gravity of Mars. The base of the mush ocean would stabilize at ~150 km/0.4 = 375 km depth. The 'room' problem for basaltic crust to subduct into the mush is less severe than on the Earth. However, quantitative geodynamics is not ready for extraterrestrial export to the early active history of Mars and Venus. Rather with exploration, geodynamics is likely to import calibration from the excellent geologic record of Mars. Application to extrasolar planets will likely remain problematic until direct observations on their tectonics become available (Lenardic and Crowley, 2012).

The Cretaceous–Tertiary asteroid impact linked geology, astronomy, and biology. The conditions following an ocean-boiling impact are even more foreign to the peaceful environments studied by traditional paleontology. Geology now provides a shopping list of feasible conditions for the early Earth. Molecular biology provides eyewitness accounts in the genomes of extant organisms. It has the potential to resurrect information lost from our rock record. Conversely, biology has had a significant effect on the crust, its tectonics, and even the mantle over geologic time.

Appendix A Thermal Models

Figures 1–3 represent generalizable features of the Earth's thermal evolution. The calculations represent realizable cases. There is no effort to tweak the model parameters other than adjusting so that the current potential temperature is ~1350 °C and the current heat flow is ~0.07 W m^{-2}. Petrologists do much better inferring the relative variations of temperature than they do with the absolute temperature (Herzberg et al., 2007; Putirka, 2005). The current ambient temperature 1350 °C maintains consistency with the work of Herzberg et al. (2010). Heat flow is in W m^{-2} and age B is billion years ago in the following equations.

The radioactive heat production is

$$q_{\text{rad}} = 0.035 \cdot 2^{B/2.5} \qquad [A.1]$$

The rate of temperature change per billion years is 180 K BY^{-1} at a net heat flow of 0.07 W m^{-2}. Mush ocean heat flow is

$$q = 0.3 \exp\left[\frac{T - 1550\,^{\circ}\text{C}}{150\,\text{K}}\right] \qquad [A.2]$$

This corresponds to T_η = 50 K. Plate tectonic heat flow is the maximum of

$$q = 0.0934 - 0.0234\left[\frac{T - 1450\,^{\circ}\text{C}}{100\,\text{K}}\right]^2 \qquad [A.3]$$

and 0.02 W m^{-2}. This provides a weak temperature dependence with a maximum at 1450 °C. The extra digits tweaked the model to the assumed current conditions. The stagnant-lid heat flow is

$$q = 0.02 \exp\left[\frac{T - 1350\,^{\circ}\text{C}}{150\,\text{K}}\right] \qquad [A.4]$$

This also corresponds to T_η = 50 K. The branch jumps on cooling are 1550 °C for mush ocean to plates and 1300 °C for plates to stagnant lid. The monotonic model remained at the transition temperature until radioactive heat generation was lower than the heat flow of the cooler mode. The nonmonotonic model heated up in the plate mode and jumped to mush ocean at 1650 °C. No jump back to plates occurred in this model, as the interior did not heat up much.

Appendix B Convection Beneath Free-Slip Lid

As envisioned, solid-state convection within the underlying mantle was the process that limited heat flow through the mush ocean. Magma entering the mush ocean froze. The temperature at the base of the mush ocean was thus less than that of the entering magma and of the ambient mantle. The difference between this temperature and the temperature of the ambient mantle drove mantle convection. The difference was probably a few hundred kelvins. I show in this appendix that the heat flow was insensitive to this rheological temperature contrast ΔT_M. I define the dimensionless variable $A \equiv \Delta T_M / T_\eta$ to compact notation.

I use the numerically simple situation of convection beneath a conductive layer with a free-slip bottom as a proxy for a complicated mush chamber (Figure B.1). I discuss the numerical method at the end of this appendix. From eqn [8], the heat flow should scale to $A^{4/3}$ when the temperature contrast is smaller than 1. The stagnant-lid heat flow in eqn [8] provides a natural scaling. This normalized heat flow may be greater than 1 as the upper part of the rheological boundary layer flows efficiently. The eyeball curve in the form of the linear rheology part of eqn [18] is

$$\left[\frac{q_{\text{CL}}}{q_{\text{SL}}}\right]^{3/4} = \left[\frac{A}{3}\right] \exp[(4.5 - A)/3.8] \qquad [B.1]$$

The heat flow reaches a maximum at a normalized rheological temperature contrast of ~3.5. The transition to ordinary stagnant-lid convection occurs at higher values. Solomatov and Moresi (2000) gave the transition at a normalized contrast of 8 for convection beneath a free-slip isothermal boundary. The

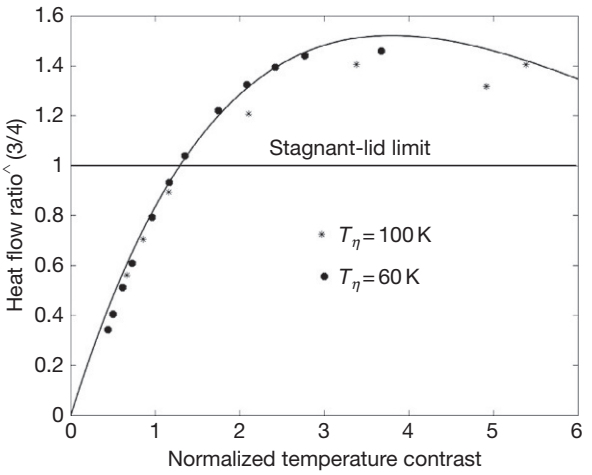


Figure B.1 Numerical results for convection beneath a free-slip conducting lid in normalized heat flow and normal rheological temperature contrast A. I varied the rheological temperature contrast to show that the scaling relationship applies.

curve [B.1] extrapolates to 1 at 8.5, compatible with this constraint.

The heat flow is slightly greater than the stagnant-lid limit over a broad range of rheological temperature contrast. One thus does not have to know the details of the gabbroic mush to compute the stagnant-lid heat flow. One does have to add the contribution of subduction of dense basaltic slabs into the mantle. Both the mush ocean parameterization [A.1] and the stagnant-lid parameterization [A.3] have the form of eqn [8] with a ratio of 3.95, rather than ~1.5 in eqn [B.1]. That is, the model implicitly includes a significant slab contribution in the parameterized model.

I use two-dimensional thermal models to find the scaling relationship [B.1] using the stream function code by Sleep (2002, 2003, 2005). It imposes a conducting lid from 197.5 km depth to the surface where it was 0 °C. The models have a 5 km grid. It defines temperature at integers times 5 km points in depth and horizontal coordinate. It defines the stream function at integer times 5 km plus 2.5 km points. This boundary condition naturally applies at stream function points. It has free-slip boundary conditions where the vertical velocity was zero and the horizontal shear traction zero to represent a lubricated base of the conducting lid.

The domain of the calculation is 900 km wide. It has free-slip mechanical boundary conditions at each end and no horizontal heat flow as a thermal boundary condition. The permeable bottom boundary condition at 500 km depth represents a large underlying adiabatic region. Fluid enters the domain at the potential temperature of 1300 °C. There is constant pressure at the boundary so that it does no work on the domain of the model. There is also no horizontal velocity at the boundary.

Well-constrained parameters are used in these generic models. Parameters include the thermal expansion coefficient $\alpha = 3 \times 10^{-5}\ \mathrm{K}^{-1}$, volume-specific heat $\rho C = 4 \times 10^{6}\ \mathrm{J\ m^{-3}\ K^{-1}}$, thermal conductivity $k = 3\ \mathrm{W\ m^{-1}\ K^{-1}}$, the potential temperature of the mantle adiabat $T_L = 1300\ ^\circ\mathrm{C}$, the density $\rho = 3400\ \mathrm{kg\ m^{-3}}$, and the acceleration of gravity $9.8\ \mathrm{m\ s^{-2}}$.

This situation is also relevant to upper boundary layers in a stratified magma chamber and convection within D″ if it is chemically denser and more viscous than the overlying mantle.

Acknowledgments

Pat Hurley introduced me to the question of the evolution of tectonics over the Earth's history while I was a graduate student at MIT. Brian Windley showed me that there are enough observations to address this issue in the deep past. I thank Kevin Zahnle for reviewing a draft manuscript of the previous version and numerous helpful comments. Euan Nisbet pointed out the existence of ancient detrital diamonds. Amy Weislogel pointed out and provided her detrital zircon data. David Stevenson, Nick Arndt, Kevin Zahnle, Euan Nisbet, Theodore Bornhorst, Laura Schaefer, and William Cassata quickly responded to queries relevant to the revision. This research was in part supported by NSF grants EAR-0406658 and EAR-0909319. The latter grant is funded under the American Recovery and Reinvestment Act of 2009 (ARRA) (Public Law 111–5). This work was performed as part of collaboration with the NASA Astrobiology Institute Virtual Planetary Laboratory Lead Team.

References

Abbott DL, Burgess L, Longhi J, and Smith WHF (1994) An empirical thermal history of the Earth's upper mantle. *Journal of Geophysical Research* 99: 13835–13850.

Abbott SS, Harrison MT, Schmitt AK, and Mojzsis ST (2012) A search for thermal excursions from ancient extraterrestrial impacts using Hadean zircon Ti–U–Th–Pb depth profiles. *Proceedings of the National Academy of Sciences* 109(34): 13486–13492.

Abe Y, Abe-Ouchi A, Sleep NH, and Zahnle KJ (2011) Habitable zone limits for dry planets. *Astrobiology* 11(5): 443–460. http://dx.doi.org/10.1089/ast.2010.0545.

Adam J, Rushmer T, O'Neil J, and Francis D (2012) Hadean greenstones from the Nuvvuagittuq fold belt and the origin of the Earth's early continental crust. *Geology* 40(4): 363–366. http://dx.doi.org/10.1130/G32623.1.

Ahrens TJ and O'Keefe JD (1972) Shock melting and vaporization of lunar rocks and minerals. *Moon* 4: 214–249.

Andrault D, Bolfan-Casanova N, Lo Nigro G, Bouhifd M, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259.

Arai S and Ishimaru S (2011) Zincian chromite inclusions in diamonds: Evidence of deep recycling origin. *Journal of Mineralogical and Petrological Sciences* 106: 85–90.

Araki T, et al. (2005) Experimental investigation of geologically produced antineutrinos with KamLAND. *Nature* 436: 499–503.

Arndt N (2003) Komatiites, kimberlites, and boninites. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB002157.

Arndt NT and Nisbet EG (2012) Processes on the young Earth and the habitats of early life. *Annual Review of Earth and Planetary Sciences* 40: 521–549.

Atwater T (1970) Implications of plate tectonics for Cenozoic tectonic evolution of western North America. *Geological Society of America Bulletin* 81: 3513–3536.

Bédard JH, Brouillette P, Madore L, and Berclaz A (2003) Archaean cratonization and deformation in the northern Superior Province, Canada: An evaluation of plate tectonic versus vertical tectonic models. *Precambrian Research* 127: 61–87.

Bédard JH, Harris LB, and Thurston PC (2012) The hunting of the snArc. *Precambrian Research* 229: 20–48.

Bell EA, Harrison TM, McCulloch MT, and Young, ED (2011) Early Archean crustal evolution of the Jack Hills Zircon source terrane inferred from Lu–Hf, $^{207}Pb/^{206}Pb$, and $\delta^{18}O$ systematics of Jack Hills zircons. *Geochimica et Cosmochimica Acta* 75(17): 4816–4829.

Blake TS, Buick R, Brown SJA, and Barley ME (2004) Geochronology of a Late Archaean flood basalt province in the Pilbara Craton, Australia: Constraints on basin evolution, volcanic and sedimentary accumulation, and continental drift rates. *Precambrian Research* 133: 143–173.

Bleeker W (2003) The late Archean record: A puzzle in ca. 35 pieces. *Lithos* 71: 99–134.

Bottke WF, et al. (2012) An Archaean heavy bombardment from a destabilized extension of the asteroid belt. *Nature* 485: 78–81.

Boussau B and Gouy M (2012) What genomes have to say about the evolution of the Earth. *Gondwana Research* 21: 483–494.

Bradley DC (2011) Secular trends in the geologic record and the supercontinent cycle. *Earth-Science Reviews* 108: 16–33.

Brown M (2006) A duality of metamorphic styles is the hallmark of plate tectonics. *Geology* 34: 961–964.

Burke K and Kidd WSF (1978) Were Archean continental geothermal gradients much steeper than those of today? *Nature* 272: 240–241.

Canil D (2004) Mildly incompatible elements in peridotites and the origins of mantle lithosphere. *Lithos* 77: 375–393.

Cassata WS, Renne PR, and Shuster DL (2011) Argon diffusion in pyroxenes: Implications for thermochronometry and mantle degassing. *Earth and Planetary Science Letters* 304: 407–416.

Cavosie AJ, Wilde SA, Liu D, Weiblen PW, and Valley JW (2004) Internal zoning and U–Th–Pb chemistry of Jack Hills detrital zircons: A mineral record of early Archean to Mesoproterozoic (4348–1576 Ma) magmatism. *Precambrian Research* 135: 251–279.

Chen SF, Witt WK, and Liu SF (2001) Transpression and restraining jogs in the northeastern Yilgarn craton, Western Australia. *Precambrian Research* 106: 309–328.

Coffin MF, Pringle MS, Duncan RA, et al. (2002) Kerguelen hotspot output since 130 Ma. *Journal of Petrology* 43(7): 1121–1139.

Condie KC (2008) Did the character of subduction change at the end of the Archean? Constraints from convergent-margin granitoids. *Geology* 36(8): 611–614.
Condie KC and Benn K (2006) Archean geodynamics: Similar to or different from modern geodynamics. In: Benn K, Mareschal J-C, and Condie KC (eds.) *Archean Geodynamics and Environments*. *Geophysical Monograph Series*, vol. 164, pp. 47–59. Washington, DC: American Geophysical Union.
Crameri F, Tackley PJ, Meilick I, Gerya TV, and Kaus BJP (2012) A free plate surface and weak oceanic crust produce single-sided subduction on Earth. *Geophysical Research Letters* 39: L03306. http://dx.doi.org/10.1029/2011GL050046.
Davaille A and Jaupart C (1993a) Thermal convection in lava lakes. *Geophysical Research Letters* 20: 1827–1830.
Davaille A and Jaupart C (1993b) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
Davaille A and Jaupart C (1994) The onset of thermal convection in fluids with temperature-dependent viscosity: Application to the oceanic mantle. *Journal of Geophysical Research* 99: 19853–19866.
Davies GF (1992) On the emergence of plate tectonics. *Geology* 20: 963–966.
Davies GF (1995) Punctuated tectonic evolution of the earth. *Earth and Planetary Science Letters* 136: 363–369.
Davies GF (2006) Gravitational depletion of the early Earth's upper mantle and the viability of early plate tectonics. *Earth and Planetary Science Letters* 243: 376–382.
Davies GF (2008) Episodic layering of the early mantle by the 'basalt barrier' mechanism. *Earth and Planetary Science Letters* 75: 382–392.
de Ronde CEJ and de Wit MJ (1994) Tectonic history of the Barberton greenstone belt, South Africa: 490 million years of Archean crustal evolution. *Tectonics* 13: 983–1005.
de Ronde CEJ and Kamo SL (2000) An Archean arc-arc collision event: A short-lived (Ca 3 Myr) episode, Weltevreden area, Barberton greenstone belt, South Africa. *Journal of African Earth Sciences* 30: 219–248.
de Wit MJ (2004) Archean greenstone belts do contain fragments of ophiolites. In: Kusky TM (ed.) *Precambrian Ophiolites and Related Rocks*. *Developments in Precambrian Geology*, vol. 13, pp. 599–614. London: Elsevier. http://dx.doi.org/10.1016/S0166-2635(04)13018-1.
Dick HJB, Lin J, and Schouten H (2003) An ultraslow-spreading class of ocean ridge. *Nature* 426: 405–412.
Dobson DP and Brodholt JP (2005) Subducted banded iron formations as a source of ultralow-velocity zones at the core-mantle boundary. *Nature* 434: 371–374.
Doin M-P, Fleitout L, and Christensen U (1997) Mantle convection and stability of depleted and undepleted continental lithosphere. *Journal of Geophysical Research* 102: 2771–2787.
Dunn SJ, Nemchin AA, Cawood PA, and Pidgeon RT (2005) Provenance record of the Jack Hills metasedimentary belt: Source of the Earth's oldest zircons. *Precambrian Research* 138: 235–254.
Dye ST (2012) Geo-neutrinos and the radioactive power of the Earth. *Reviews of Geophysics* 50(3). (arXiv:1111.6099v2).
England P and Bickle M (1984) Continental thermal and tectonic regimes during the Archean. *Journal of Geology* 92: 353–367.
Enomoto S, Ohtani E, Inoue K, and Suzuki A (2007) Neutrino geophysics with KamLAND and future prospects. *Earth and Planetary Science Letters* 258: 147–159.
Erickson EJ (2010) Structural and kinematic analysis of the Shagawa Lake shear zone, Superior Province, northern Minnesota: Implications for the role of vertical versus horizontal tectonics in the Archean. *Canadian Journal of Earth Sciences* 47: 1463–1479.
Farquhar J, Wing BA, McKeegan KD, Harris JW, Cartigny P, and Thiemens MH (2002) Mass-independent sulfur of inclusions in diamond and sulfur recycling on the early Earth. *Science* 298: 2369–2372.
Fiorentini G, Lissia M, Mantovani F, and Vannucci R (2005) Geo-neutrinos: A new probe of earth's interior. *Earth and Planetary Science Letters* 238: 235–247.
Foley SF (2011) A reappraisal of redox melting in the Earth's mantle as a function of tectonic setting and time. *Journal of Petrology* 52: 1363–1391.
Fontaine FJ, Olive J-A, Cannat M, Escartin J, and Perol T (2011) Hydrothermally induced melt lens cooling and segmentation along the axis of fast and intermediate-spreading centers. *Geophysical Research Letters* 38: L14307. http://dx.doi.org/10.1029/2011GL047798.
Friend CRL, Bennett VC, and Nutman AP (2002) Abyssal peridotites >3,800 Ma from southern West Greenland: Field relationships, petrography, geochronology, whole-rock and mineral chemistry of dunite and harzburgite inclusions in the Itsaq Gneiss Complex. *Contributions to Mineralogy and Petrology* 143: 71–92.
Friend CRL and Nutman AP (2010) Eoarchean ophiolites? New evidence for the debate on the Isua supracrustal belt, southern west Greenland. *American Journal of Science* 310: 826–861. http://dx.doi.org/10.2475/09.2010.04.
Furnes H, de Wit M, Staudigel H, Rosing M, and Muehlenbachs K (2007) A vestige of Earth's oldest ophiolite. *Science* 315: 1704–1707. http://dx.doi.org/10.1126/science.1139170.
Furnes H, Rosing M, Dilek Y, and de Wit M (2009) Isua supracrustal belt (Greenland) – A vestige of a 3.8 Ga suprasubduction zone ophiolite, and the implications for Archean geology. *Lithos* 113: 115–132. http://dx.doi.org/10.1016/j.lithos.2009.03.043.
Galer SJG and Mezger K (1998) Metamorphism denudation and sea level in the Archean and cooling of the Earth. *Precambrian Research* 92: 387–412.
Goldblatt C, Zahnle KJ, Sleep NH, and Nisbet EG (2009) The Eons of Chaos and Hades. *Solid Earth Discussion* 1: 47–53.
Grassineau NV, Abell PWU, Fowler CMR, and Nisbet EG (2005) Distinguishing biological from hydrothermal signatures via sulphur and carbon isotopes in Archaean mineralizations at 3.8 and 2.7 Ga. In: McDonald I, Boyce AL, Butler IB, Herrington RJ, and Polya DA (eds.) *Mineral Deposits and Earth Evolution*, vol. 248, pp. 195–212. London: Geological Society, Special Publication.
Grocholski B, Catalli K, Shim S-H, and Prakapenha V (2012) Mineralogical effects on the detectability of the postperovskite boundary. *Proceedings of the National Academy of Sciences* 109(7): 2275–2279.
Grove TL and Parman SW (2004) Thermal evolution of the Earth as recorded by komatiites. *Earth and Planetary Science Letters* 219: 173–187.
Harrison TM (2009) The Hadean crust: Evidence from >4 Ga zircons. *Annual Review of Earth and Planetary Sciences* 37: 479–505.
Harrison TM, Blichert-Toft J, Muller W, Albarede F, Holden P, and Mojzsis SJ (2005) Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* 310: 1947–1950.
Hastie AR, Kerr AC, McDonald I, et al. (2010a) Geochronology, geochemistry and petrogenesis of rhyodacite lavas in eastern Jamaica: A new adakite subgroup analogous to early Archaean continental crust? *Chemical Geology* 276: 344–359.
Hastie AR, Kerr AC, McDonald I, et al. (2010b) Do Cenozoic analogues support a plate tectonic origin for Earth's earliest continental crust? *Geology* 38: 495–498.
Hazen RM, Papineau D, Bleeker W, et al. (2008) Mineral evolution. *American Mineralogist* 93: 1693–1720.
Herzberg C, Asimow PD, Arndt N, et al. (2007) Temperatures in ambient mantle and plumes: Constraints from basalts, picrites and komatiites. *Geochemistry, Geophysics, Geosystems* 8: Q02006. http://dx.doi.org/10.1029/2006GC001390.
Herzberg C, Condie K, and Korenaga J (2010) Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters* 292: 79–88.
Herzberg C and Rudnick R (2012) Formation of cratonic lithosphere: An integrated thermal and petrological model. *Lithos* 149: 4–15.
Hickman AH (2004) Two contrasting granite-greenstone terranes in the Pilbara Craton, Australia: Evidence for vertical and horizontal tectonic regimes prior to 2900. *Precambrian Research* 131: 153–172.
Hirschmann MM (2012) Magma ocean influence on early atmosphere mass and composition. *Earth and Planetary Science Letters* 341–344: 48–57.
Hoffmann JE, Münker C, Næraa T, et al. (2011) Mechanisms of Archean crust formation inferred from high-precision HFSE systematics in TTGs. *Geochimica et Cosmochimica Acta* 75: 4157–4178.
Hofmann A and Kusky TM (2004) The Belingwe Greenstone belt: Ensialic or oceanic? In: Kusky TM (ed.) *Precambrian Ophiolites and Related Rocks*. *Developments in Precambrian Geology*, vol. 13, pp. 487–538. Amsterdam: Elsevier.
Hopkins MD, Harrison TM, and Manning CE (2010) Constraints on Hadean geodynamics from mineral inclusions in >4 Ga zircons. *Earth and Planetary Science Letters* 298: 367–376. http://dx.doi.org/10.1016/j.epsl.2010.08.010.
Huebeck C and Lowe DR (1994) Late syndepositional deformation and detachment tectonics in the Barberton Greenstone Belt, South Africa. *Tectonics* 13: 1514–1536.
Hurley PM and Rand JR (1969) Pre-drift continental nuclei. *Science* 164: 1229–1242.
Iizuka T, Nakai S, Sahoo YV, Takamasa A, Hirata T, and Maruyama S (2010) The tungsten isotopic composition of Eoarchean rocks: Implications for early silicate differentiation and core–mantle interaction on Earth. *Earth and Planetary Science Letters* 291: 189–200.
Ingle S, Weis D, and Frey FA (2002) Indian continental crust recovered from Elan Bank, Kerguelen Plateau (ODP Leg 183, Site 1137). *Journal of Petrology* 43(7): 1241–1257.
Jackson MG and Carlson RW (2012) Homogeneous superchondritic $^{142}Nd/^{144}Nd$ in the mid-ocean ridge basalt and ocean island basalt mantle. *Geochemistry, Geophysics, Geosystems* 13: Q06011. http://dx.doi.org/10.1029/2012GC004114.
Jacob DE, Schmickler B, and Schulze DJ (2003) Trace element geochemistry of coesite-bearing eclogites from the Roberts Victor kimberlite, Kaapvaal craton. *Lithos* 71: 337–351.
Johnson BC and Melosh HJ (2012) Impact spherules as a record of an ancient heavy bombardment of Earth. *Nature* 485: 75–77.

Kasting JF and Catling D (2003) Evolution of a habitable planet. *Annual Review of Astronomy and Astrophysics* 41: 429–463.

Kemp AIS, Wilde SA, Hawkesworth CJ, et al. (2010) Hadean crustal evolution revisited: New constraints from Pb–Hf isotope systematics of the Jack Hills zircons. *Earth and Planetary Science Letters* 296: 45–56.

Kinoshita N, et al. (2012) A shorter ^{146}Sm half-life measured and implications for ^{146}Sm–^{142}Nd chronology in the Solar System. *Science* 335: 1614–1617.

Kisters AFM, Stevens G, Dziggel A, and Armstrong RA (2003) Extensional detachment faulting and core-complex formation in the southern Barberton granite-greenstone terrain, South Africa: Evidence for a 3.2 Ga orogenic collapse. *Precambrian Research* 127: 355–378.

Kitajima K, Maruyama S, Utsunomiya S, and Liou JG (2001) Seafloor hydrothermal alteration at an Archaean mid-ocean ridge. *Journal of Metamorphic Geology* 19: 581–597.

Klein EM (2003) Geochemistry of the igneous oceanic crust. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 3 (edited by RL Rudnick), ch. 3,13, Oxford: Elsevier–Pergamon. pp. 433–463.

Klein EM and Langmuir CH (1987) Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *Journal of Geophysical Research* 92: 8089–8115.

Kopylova MG, Afanasiev VP, Bruce LF, Thurston P, and Ryder J (2011) Metaconglomerate preserves evidence for kimberlite, diamondiferous root and medium grade terrane of a pre-2.7 Ga Southern Superior protocraton. *Earth and Planetary Science Letters* 312: 213–225.

Korenaga J (2006) Archean geodynamics and the thermal evolution of the Earth. In: Benn K, Mareschal J-C, and Condie KC (eds.) *Archean Geodynamics and Environments*. *Geophysical Monograph Series*, vol. 164, pp. 7–32. Washington, DC: American Geophysical Union.

Korenaga J (2008a) Plate tectonics, flood basalts, and the evolution of Earth's oceans. *Terra Nova* 20: 419–439.

Korenaga J (2008b) Urey ratio and the structure and evolution of Earth's mantle. *Reviews of Geophysics* 46: RG2007. http://dx.doi.org/10.1029/2007RG000241.

Korenaga J (2011) Thermal evolution with a hydrating mantle and the initiation of plate tectonics in the early Earth. *Journal of Geophysical Research* 116: B12403. http://dx.doi.org/10.1029/2011JB008410.

Kuiper YD, Lin S, and Böhm CO (2011) Himalayan-type escape tectonics along the Superior Boundary Zone in Manitoba, Canada. *Precambrian Research* 197: 248–262.

Kusky TM and Li JH (2010) Origin and emplacement of Archean ophiolites of the central orogenic belt, North China craton. *Journal of Earth Science* 21(5): 744–781. http://dx.doi.org/10.1007/s12583-010-0119-8.

Kusky TM, Li JH, and Tucker RD (2001) The Archean Dongwanzi ophiolite complex, North China craton: 2.505-billion-year-old oceanic crust and mantle. *Science* 292: 1142–1145.

Kusky TM and Zhai M (2012) The Neoarchean Ophiolite in the North China Craton: Early Precambrian plate tectonics and scientific debate. *Journal of Earth Science* 23(3): 277–284. http://dx.doi.org/10.1007/s12583-012-0253-6.

Kvande H, Linga H, Motzfeldt K, and Wahlbeck PK (1979) A thermogravimetric method for determination of vapour pressures above 10^{-2} atm. II. Vapour pressure of molten sodium chloride. *Acta Chemica Scandinavica* A33: 281–288.

Lane N, Allen JF, and Martin W (2010) How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays* 32: 271–280.

Lenardic A and Crowley JW (2012) On the notion of well-defined tectonic regimes for terrestrial planets in this solar system and others. *Astrophysical Journal* 755: 132. http://dx.doi.org/10.1088/0004-637X/755/2/132.

Lenardic A, Moresi LN, Jellinek AM, and Manga M (2005) Continental insulation, mantle cooling, and the surface area of oceans and continents. *Earth and Planetary Science Letters* 234: 317–333.

Lenardic A, Moresi L-N, and Mühlhaus H (2003) Longevity and stability of cratonic lithosphere: Insights from numerical simulations of coupled mantle convection and continental tectonics. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001859.

Liebske C, Schmickler B, Terasaki H, et al. (2005) Viscosity of peridotite liquid up to 13 GPa: Implications for magma ocean viscosities. *Earth and Planetary Science Letters* 240: 589–604.

Lin S (2005) Synchronous vertical and horizontal tectonism in the Neoarchean: Kinematic evidence from a synclinal keel in the northwestern Superior craton, Canada. *Precambrian Research* 139: 181–194.

Lyubetskaya TJ and Korenaga J (2007) Chemical composition of Earth's primitive mantle and its variance: 2. Implications for global geodynamics. Journal of Geophysical Research 112: B03212, doi: http://dx.doi.org/10.1029/2005JB004224.

Mareschal J-C, Jaupart C, Phaneuf C, and Perry C (2012) Geoneutrinos and the energy budget of the Earth. *Journal of Geodynamics* 54: 43–54. http://dx.doi.org/10.1016/j.jog.2011.10.005.

McDonough WF and Sun S (1995) The composition of the Earth. *Chemical Geology* 120: 223–253. http://dx.doi.org/10.1016/0009-2541(94)00140-4.

McGuire J and Beroza G (2012) A rogue earthquake off Sumatra. *Science* 336: 1118–1119.

McKenzie D and Bickle MJ (1988) The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology* 29: 625–679.

Menzies AH, Carlson RW, Shirey SB, and Gurney JJ (2003) Re–Os systematics of diamond-bearing eclogites from the Newlands kimberlite. *Lithos* 71: 323–336.

Molnar P and Tapponnier P (1975) Cenozoic tectonics of Asia: Effects of a continental collision. *Science* 189: 419–425.

Moores EM (2002) Pre-1 Ga (pre Rodinian) ophiolites: Their tectonic and environmental implications. *Geological Society of America Bulletin* 114: 80–95.

Morgan JK, Moore GF, and Clague DA (2003) Slope failure and volcanic spreading along the submarine south flank of Kilauea volcano, Hawaii. *Journal of Geophysical Research* 108: 2415. http://dx.doi.org/10.1029/2003JB002411.

Moyen J-F (2009) High Sr/Y and La/Yb ratios: The meaning of the "adakitic signature". *Lithos* 112: 556–574.

Mulkidjanian AV, Bychkov AYu, Dibrova DV, Galperin MY, and Koonin EV (2012) Origin of first cells at terrestrial, anoxic geothermal fields. *Proceedings of the National Academy of Sciences* 109(14): E821–E830.

Næraa T, Scherstén A, Rosing MT, et al. (2012) Hafnium isotope evidence for a transition in the dynamics of continental growth 3.2 Gyr ago. *Nature* 485: 627–631.

Nagel TJ, Hoffmann JE, and Münker C (2012) Generation of Eoarchean tonalite-trondhjemite-granodiorite series from thickened mafic arc crust. *Geology* 40(4): 375–378.

Nakamura K and Kato Y (2004) Carbonatization of oceanic crust by the seafloor hydrothermal activity and its significance as a CO_2 sink in the Early Archean. *Geochimica et Cosmochimica Acta* 68: 4595–4618.

Nisbet EG (1991) *Living Earth: A Short History of Life and Its Home*. London: Harper Collins Academic, 237 pp.

Okubo PG, Benz HM, and Chouet BA (1997) Imaging the crustal magma sources beneath Mauna Loa and Kilauea volcanoes, Hawaii. *Geology* 25: 867–870.

O'Neil J, Carlson RW, Francis D, and Stevenson RK (2008) Neodymium-142 evidence for Hadean mafic crust. *Science* 321: 1828–1831. http://dx.doi.org/10.1126/science.1161925.

O'Neil J, Carlson RW, Paquette J-L, and Francis D (2012) Formation age and metamorphic history of the Nuvvuagittuq Greenstone Belt. *Precambrian Research* 220–221: 23–44.

O'Neil J, Francis D, and Carlson RW (2011) Implications of the Nuvvuagittuq greenstone belt for the formation of Earth's early crust. *Journal of Petrology* 52: 985–1009. http://dx.doi.org/10.1093/petrology/egr014.

O'Neill C, Lenardic A, Moresi L-N, Torsvik T, and Lee CA (2007) Episodic Precambrian subduction. *Earth and Planetary Science Letters* 262: 552–562.

O'Neill C and Wyman DA (2006) Geodynamic modeling of Late Archean subduction: Pressure–temperature constraints from greenstone belt diamond deposits. In: Benn K, Mareschal J-C, and Condie KC (eds.) *Archean Geodynamics and Environments*. *Geophysical Monograph Series*, vol. 164, pp. 177–188. Washington, DC: American Geophysical Union.

Padhi CM, Korenaga J, and Ozima M (2012) Thermal evolution of Earth with xenon degassing: A self-consistent approach. *Earth and Planetary Science Letters* 341–344: 1–9.

Papineau D (2010) Mineral environments on the earliest Earth. *Elements* 6: 25–30.

Parmentier AC, Lin S, and Cockery MT (2006) Structural evolution of the Cross Lake greenstone belt in the northwestern Superior Province, Manitoba: Implications for relationship between vertical and horizontal tectonism. *Canadian Journal of Earth Sciences* 43(7): 767–787. http://dx.doi.org/10.1139/E06-006.

Pavlov A and Kasting JF (2002) Mass-independent fractionation of sulfur isotopes in Archean sediments: Strong evidence for an anoxic Archean atmosphere. *Astrobiology* 2: 27–41.

Plank T and Langmuir CH (1992) Effects of the melting regime on the composition of the oceanic crust. *Journal of Geophysical Research* 97: 19749–19770.

Plank T and Langmuir CH (1998) The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chemical Geology* 145(3): 325–394.

Polat A and Kerrich R (2006) Reading the geochemical fingerprints of Archean hot subduction volcanic rocks: Evidence for accretion and crustal recycling in a mobile tectonic regime. In: Benn K, Mareschal J-C, and Condie KC (eds.) *Archean Geodynamics and Environments*. *Geophysical Monograph Series*, vol. 164, pp. 189–213. Washington, DC: American Geophysical Union.

Pons M-L, Quitté G, Fujii T, et al. (2011) Early Archean serpentine mud volcanoes at Isua, Greenland, as a niche for early life. *Proceedings of the National Academy of Sciences of the United States of America* 108: 17639–17643.
Potter RWK, Collins GS, Kiefer WS, McGovern PJ, and Kring DA (2012) Constraining the size of the South Pole-Aitken basin impact. *Icarus* 220: 730–743.
Putirka KD (2005) Mantle potential temperatures at Hawaii, Iceland, and the mid-ocean ridge system, as inferred from olivine phenocrysts: Evidence for thermally driven mantle plumes. *Geochemistry, Geophysics, Geosystems* 6: Q05L08.
Ritzwoller MH (1978–2012) Shapiro NM, and Leahy GM (2003) A resolved mantle anomaly as the cause of the Australian-Antarctic Discordance. *Journal of Geophysical Research: Solid Earth* 108: B12.
Rizo H, Boyet M, Blichert-Toft J, and Rosing M (2011) Combined Nd and Hf isotope evidence for deep-seated source of Isua lavas. *Earth and Planetary Science Letters* 312: 267–279.
Rosing MT, Bird DK, Sleep NH, Glassley W, and Albarede F (2006) The rise of continents – An essay on the geologic consequences of photosynthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology* 232(2–4): 99–113.
Roth A, Scherer E, Maden C, Mezger K, and Bourdon B (2013) Revisiting the ^{142}Nd deficits in the 1.48 Ga Khariar alkaline rocks, India. *Chemical Geology 386:238–248*.
Rüpke LH, Morgan JP, Hort M, and Connolly JAD (2004) Serpentine and the subduction zone water cycle. *Earth and Planetary Science Letters* 223: 17–34.
Russell MJ, Hall AJ, and Martin W (2010) Serpentinization as a source of energy at the origin of life. *Geobiology* 8: 355–371.
Ryder G (2002) Mass flux in the ancient Earth-Moon system and benign implications for the origin of life on Earth. *Journal of Geophysical Research* 107(E4): 5022.
Saltzer RL, Chatterjee N, and Grove TL (2001) The spatial distribution of garnets and pyroxenes in mantle peridotites: Pressure–temperature history of peridotites from the Kaapvaal craton. *Journal of Petrology* 42: 2229–2231.
Saura E, Vergés J, Brown D, et al. (2008) Structural and tectonic evolution of western Cuba fold and thrust belt. *Tectonics* 27: TC4002.
Schaefer L, Lodders K, and Fegley B Jr. (2012) Vaporization of the Earth: Application to exoplanet atmospheres. *Astrophysical Journal* 755: 41. http://dx.doi.org/10.1088/0004-637X/755/1/41.
Schulze DJ, Harte B, Valley JW, Brenan JM, and Channer DMde R (2003a) Extreme crustal oxygen isotope signatures preserved in coesite in diamond. *Nature* 423: 68–70.
Schulze DJ, Valley JW, Spicuzza MJ, and Channer DMde R (2003b) The oxygen isotope composition of eclogitic and peridotitic garnet xenoliths from the La Ceniza kimberlite, Guaniamo, Venezuela. *International Geology Review* 45: 968–975.
Serrano-Andrés L and Merchán M (2009) Are the five natural DNA/RNA base monomers a good choice from natural selection? A photochemical perspective. *Journal of Photochemistry and Photobiology, C: Photochemistry Reviews* 10: 21–32.
Shirey SB and Richardson SH (2011) Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* 333: 434–436.
Sibley DF and Vogel TA (1976) Chemical mass balance of the Earth's crust: The calcium dilemma and the role of pelagic sediments. *Science* 192: 551–553.
Silver PG and Behn MD (2008) Intermittent plate tectonics? *Science* 319: 85–88.
Sinton JM and Detrick RS (1992) Mid-ocean ridge magma chambers. *Journal of Geophysical Research* 97: 197–216.
Sizova E, Gerya T, Brown M, and Perchuk LL (2010) Subduction styles in the Precambrian: Insight from numerical experiments. *Lithos* 116: 209–229.
Slack JF and Cannon WF (2009) Extraterrestrial demise of banded iron formations 1.85 billion years ago. *Geology* 11: 1011–1014.
Sleep NH (1979) Thermal history and degassing of the Earth: Some simple calculations. *Journal of Geology* 87: 671–686.
Sleep NH (1992) Archean plate-tectonics – What can be learned from continental geology. *Canadian Journal of Earth Sciences* 29: 2066–2071.
Sleep NH (2000) Evolution of the mode of convection within terrestrial planets. *Journal of Geophysical Research* 105: 17563–17578.
Sleep NH (2002) Local lithospheric relief associated with fracture zones and ponded plume material. *Geochemistry, Geophysics, Geosystems* 3: (2001GC000376).
Sleep NH (2003) Survival of Archean cratonal lithosphere. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2001JB000169.
Sleep NH (2005) Evolution of continental lithosphere. *Annual Review of Earth and Planetary Sciences* 33: 369–393.
Sleep NH (2010) The Hadean-Archaean environment. *Cold Spring Harbor Perspectives in Biology* 2: a002527. http://dx.doi.org/10.1101/cshperspect.a002527.
Sleep NH, Bird DK, and Pope E (2011) Serpentine and the dawn of life. *Philosophical Transactions of the Royal Society B* 366: 2857–2869. http://dx.doi.org/10.1098/rstb.2011.0129.
Sleep NH, Bird DK, and Pope E (2012) Paleontology of Earth's mantle. *Annual Review of Earth and Planetary Sciences* 40: 277–300.
Sleep NH and Windley BF (1982) Archean plate tectonics: Constraints and inferences. *Journal of Geology* 90: 363–379.
Sleep NH and Zahnle K (1998) Refugia from asteroid impacts on early Mars and the early Earth. *Journal of Geophysical Research* 103: 28529–28544.
Sleep NH and Zahnle K (2001) Carbon dioxide cycling and implications for climate on ancient Earth. *Journal of Geophysical Research* 106: 1373–1399.
Sleep NH, Zahnle K, and Neuhoff PS (2001) Initiation of clement surface conditions on the early Earth. *Proceedings of the National Academy of Sciences* 98: 3666–3672.
Sobolev AV, Hofmann AW, Sobolev SV, and Nikogosian IK (2005) An olivine-free mantle source of Hawaiian shield basalts. *Nature* 434: 590–597.
Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Solomatov VS (2004) Initiation of subduction by small-scale convection. *Journal of Geophysical Research* 109: B01412.
Solomatov VS and Moresi L-N (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on Earth and other terrestrial planets. *Journal of Geophysical Research* 105: 21795–21817.
Šrámek O, McDonough WF, and Learned JG (2012) Geoneutrinos. *Advances in High Energy Physics* 2012: 235686. http://dx.doi.org/10.1155/2012/235686..
Staudigel H (2003) Hydrothermal alteration processes in the oceanic crust. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 3 (edited by RL Rudnick), ch. 3.15, Oxford: Elsevier Pergamon. pp. 1511–535.
Stern RJ (2005) Evidence from ophiolites, blueschists, and ultrahigh-pressure metamorphic terranes that the modern episode of subduction tectonics began in Neoproterozoic time. *Geology* 33: 557–560.
Stevenson DJ (2003) Styles of mantle convection and their influence on planetary evolution. *Comptes Rendus Geoscience* 335: 99–111.
Stone WE, Crocket JH, Fleet ME, and Larson MS (1996) PGE mineralization in Archean volcanic systems: Geochemical evidence from thick, differentiated mafic-ultramafic flows, Abitibi greenstone belt, Ontario, and implications for exploration. *Journal of Geochemical Exploration* 56: 237–263.
Suganuma Y, et al. (2006) Paleomagnetism of the Marble Bar Chert Member, Western Australia: Implications for apparent polar wander path for Pilbara craton during Archean time. *Earth and Planetary Science Letters* 252: 360–371.
Tackley PJ (2000a) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations, 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1, 2000GC000036.
Tackley PJ (2000b) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations, 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1, 2000GC000043.
Tappe S, Smart KA, Pearson DG, Steenfelt A, and Simonetti A (2011) Craton formation in Late Archean subduction zones revealed by first Greenland eclogites. *Geology* 39(12): 1103–1106.
Tajika E and Matsui T (1993) Evolution of seafloor spreading rate based on ^{40}Ar degassing history. *Geophysical Research Letters* 20: 851–854.
Tessalina SG, Bourdon B, Kranendonk M, Birck JL, and Philippot P (2010) Influence of Hadean crust evident in basalts and cherts from the Pilbara Craton. *Nature Geoscience* 3: 214–217.
Thern ER and Nelson DR (2012) Detrital zircon age structure within ca. 3 Ga metasedimentary rocks, Yilgarn Craton: Elucidation of Hadean source terranes by principal component analysis. *Precambrian Research* 214–215: 28–43.
Thom R (1983) *Parabolas et Catastrophes*. Paris: Flammarion, 193 pp.
Touboul M, Puchtel I, and Walker RJ (2012) ^{182}W Evidence for long-term preservation of early mantle differentiation products. *Science* 335: 1065–1069.
Tozer DC (1970) Factors determining the temperature evolution of thermally convecting Earth models. *Physics of the Earth and Planetary Interiors* 2: 393–398.
Turcotte DL and Schubert G (2002) *Geodynamics: Applications of Continuum Physicsto Geological Problems*. New York: Wiley 456 pp.
Turner S, Rushmer T, Reagan M, and Moyen J-F (2014) Heading down early on? Start of subduction on Earth. *Geology* 42: 139–142.
Upadhyay D, Scherer EE, and Mezger K (2009) ^{142}Nd evidence for an enriched Hadean reservoir in cratonic roots. *Nature* 459: 1118–1121.
Valley JW, et al. (2014) Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nature Geoscience*. 7(3): 219–223. http://dx.doi.org/10.1038/NGEO2075.
van Hunen J and Moyen J-F (2012) Archean subduction: Fact or fiction. *Annual Review of Earth and Planetary Sciences* 40: 195–219.
Van Kranendonk MJ (2010) two types of Archean continental crust: Plume and plate tectonics on early Earth. *American Journal of Science* 310: 1187–1209. http://dx.doi.org/10.2475/10.2010.01.

Van Kranendonk MJ, Collins WJ, Hickman A, and Pawley MJ (2004) Critical tests of vertical vs. horizontal tectonic models for the Archaean East Pilbara Granite-Greenstone Terrane, Pilbara Craton, Western Australia. *Precambrian Research* 131: 173–211.

van Thienen P, Vlaar NJ, and van den Berg AP (2004) Plate tectonics on the terrestrial planets. *Physics of the Earth and Planetary Interiors* 142: 61–74.

Wallace PJ (2005) Volatiles in subduction zone magmas: Concentrations and fluxes based on melt inclusion and volcanic gas data. *Journal of Volcanology and Geothermal Research* 140: 217–240.

Walter LS and Carron MK (1964) Vapor pressure and vapor fractionation of silicate melts of tektite composition. *Geochimica et Cosmochimica Acta* 28: 937–951.

Watson EB, Thomas JB, and Cherniak DJ (2007) ^{40}Ar retention in the terrestrial planets. *Nature* 449: 299–304.

Weislogel AL, Graham SA, Chang EZ, Wooden JL, Gehrels GE, and Yang H (2006) Detrital zircon provenance of the Late Triassic Songpan-Ganzi complex: Sedimentary record of collision of the North and South China blocks. *Geology* 34(2): 97–100.

White RS, McKenzie D, and O'Nions RK (1992) Oceanic crustal thickness from seismic measurements and rare earth element inversions. *Journal of Geophysical Research* 97: 19683–19715.

Wilson DS, McCrory PA, and Stanley RG (2005) Implications of volcanism in coastal California for the Neogene deformation history of western North America. *Tectonics* 24(3): TC3008.

Yokochi R and Marty B (2005) Geochemical constraints on mantle dynamics in the Hadean. *Earth and Planetary Science Letters* 238: 17–30.

Zahnle K, Arndt N, Cockell C, et al. (2007) Emergence of a habitable planet. *Space Science Reviews* 129: 35–78.

Zahnle K and Sleep NH (2006) Impacts and the early evolution of life. In: Thomas PJ, Hicks RD, Chyba CF, and McKay CP (eds.) *Comets and the Origin and Evolution of Life*, 2nd edn., pp. 207–251. Berlin: Springer.

Zatman S, Gordon RG, and Mutnuri K (2005) Dynamics of diffuse oceanic plate boundaries: Insensitivity to rheology. *Geophysical Journal International* 162: 239–248.

Zhao G-C, Wilde SA, Li S-Z, Sun M, Grant ML, and Li X-P (2007) U–Pb zircon age constraints on the Dongwanzi ultramafic-mafic body, North China, confirm it is not an Archean ophiolite. *Earth and Planetary Science Letters* 255: 85–93. http://dx.doi.org/10.1016/j.epsl.2006.12.007.

Zoback MD and Townend J (2001) Implications of hydrostatic pore pressures and high crustal strength for the deformation of intraplate lithosphere. *Tectonophysics* 336: 19–30.

9.07 Mechanism of Continental Crustal Growth

M Stein, Geological Survey of Israel, Jerusalem, Israel
Z Ben-Avraham, Tel Aviv University, Ramat Aviv, Israel

© 2015 Elsevier B.V. All rights reserved.

9.07.1	**Introduction**	173
9.07.2	**Structure and Chemical Composition of the Continental Crust**	174
9.07.2.1	General	174
9.07.2.2	Composition of the Continental Crust: Open Questions	176
9.07.2.2.1	Problems with the simple 'andesitic model'	176
9.07.2.2.2	Plumes and arcs	176
9.07.2.2.3	The Eu dilemma	177
9.07.2.2.4	Crustal foundering	177
9.07.2.2.5	Archean TTG magmatic suites	178
9.07.2.2.6	Post-Archean granites	178
9.07.2.3	The Lithospheric Mantle and Continental Stability	178
9.07.3	**Models of Crustal Formation and Continental Growth**	179
9.07.3.1	General	179
9.07.3.2	Timing and Rates of Crustal Growth	179
9.07.3.2.1	General	179
9.07.3.2.2	Ages of the major orogens	179
9.07.3.3	The Reymer and Schubert Dilemma	180
9.07.3.4	Continental Growth and Heat Production	181
9.07.4	**Oceanic Plateau, Accreted Terrains and Juvenile Crust Additions**	181
9.07.4.1	General	181
9.07.4.2	Accretion of Allochthonous Terrains at the Pacific Margins	183
9.07.4.3	The Transformation of Oceanic Plateau to Continental Crust	185
9.07.4.4	Mantle Overturn and Crustal Formation Episodes	185
9.07.4.5	^{142}Nd and Composition of the 'Juvenile Mantle'	186
9.07.4.6	Hf in Zircons and Crust Recycling	188
9.07.5	**The Production of Calc-Alkaline and Alkaline Granites and Related Rocks**	188
9.07.5.1	General	188
9.07.5.2	Models of Calc-Alkaline Granite Production	189
9.07.5.3	The Petrogenesis of Calc-Alkaline Granites and Related Rocks from the Elat Massif	190
9.07.5.4	Crustal Anatexis and the Role of Water	190
9.07.5.5	The Genesis of Alkaline A-Type Granites	191
9.07.5.6	Petrogenesis of Granites: Concluding Comments	191
9.07.6	**Rapid Growth of Major Continental Segments (the 'Major Orogens')**	192
9.07.6.1	General	192
9.07.6.2	The ANS Orogeny (~0.9–0.6 Ga)	192
9.07.6.3	The Birimian Orogeny (~2.2–1.8 Ga)	194
9.07.6.4	The Superior Orogeny (~2.7–2.5 Ga)	194
9.07.6.5	Early Precambrian Orogenies (~2.9–2.8 Ga)	194
9.07.6.6	Early (Hadean) Crust and the Initiation of Wilsonian Plate Tectonics–MOMO Cycles	195
9.07.7	**Summary**	195
References		196

9.07.1 Introduction

The continental crust comprises the outmost 20–80 km (average thickness is about 36 km) of the solid surface of the Earth covering ~41% of the Earth surface area. Most of this area (~71%) is currently elevated above sea level, while the rest is defined by the topography of the continental shelves. The continental crust density ranges from 2.7 to 2.9 g cm^{-3} and increases with depth. The vertical extent of the continental crust is defined by the compressional seismic wave velocity that jumps from ~7 to >7.6–8 km s^{-1} across the Mohorovičić discontinuity (or 'Moho'). The Moho discontinuity discovered by the Croatian seismologist Andrija Mohorovičić is the boundary between the crust and the mantle, which reflects the different densities of the crust and the mantle, normally occurring at an 'average' depth of ~35 km beneath the continental surfaces and about 8 km beneath the oceanic basins. The original recognition and definition of this boundary

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00159-7

reflected seismic properties. However, in some regions, the Moho is not always well defined and appears to be transitional in its physical–chemical characteristics (cf. Griffin and O'Reilly, 1987).

The oceanic crust covering ~59% of the Earth surface is significantly thinner than the continental crust and is characterized by its distinct chemical composition and younger age (the oldest oceanic crust is Jurassic in age). Nevertheless, it appears that the growth of continental crust involves the accretion of thick magmatic sequences that were formed in the intraoceanic environment (e.g., oceanic plateaus); thus, the geodynamic evolution and geochemical evolution of oceanic and continental crusts are related.

The temperature gradient is typically less beneath continents than beneath oceanic basins, reflecting the thicker immobile regions underlying the continents. These regions are termed the lithospheric mantle, and they typically form a thick colder 'root' beneath ancient shields and a relatively thinner root beneath younger shields. Yet in intracontinental rifting regions, the asthenospheric mantle is rising and interacts with the lithospheric mantle and continental crust and replaces it with new mantle-derived material. Moreover, it appears that many of the current oceanic basins (and oceanic crust) commenced their history within the continental environment calling for some important relationship between stability and magmatic–thermal history of the continental lithosphere and the plate tectonic cycle. The relationship between asthenospheric mantle and lithospheric mantle and the questions concerning the growth, maturation, and fate of continental lithosphere and crust throughout geologic time are fundamental problems in Earth sciences that will be reviewed in this chapter.

The continental crust is of great antiquity and contains the record of most of the geologic (physical and chemical) evolution of the Earth. A major question in earth sciences is when and in what rates and modes the continental crust was formed and evolved. It is broadly accepted that most of the present-day continental masses are composed of rocks that were produced at the end of the Archean and the beginning of the Proterozoic, within the time period lasting from ~3.2 to ~2.0 Ga. The oldest surviving rocks were found in all continents and date from more than four billion years ago, a few hundred million years after the formation of the Earth (Taylor and McLennan, 1995). Yet although the continental crust covers about 40% of the terrestrial surface, rocks older than 2.5 Ga account for less than ~15% of the exposed continental area (Windley, 1995). Based on Nd model ages, the 'average' age of the continental crust was calculated as $\sim 2.2 \pm 0.1$ Ga, which indicates that 50–60% of the continental crust was formed before 2.6 Ga (cf. DePaolo et al., 1991; McCulloch and Bennett, 1994; Nelson and DePaolo, 1985; Taylor and McLennan, 1995). These chronological determinations are critical to the long-standing debate on the history and mechanism of crustal formation, where models of very early creation and subsequent recycling of continental crust material contrast with continuous/episodic crustal growth (the 'Armstrong–Moorbath debate' elaborated later in the text).

Much of the present-day continental creation appears to occur at convergent plate margins mainly along the subduction zones; however, several studies indicate that large amounts of continent were created in the geologic past during short 'super events' at rates that are difficult to explain through 'typical' subduction activity (**Section 9.07.3.3**). This would imply that large portions of the continents were formed by other means than convergent margin magmatism, probably processes related to hot spot or mantle plume magmatism (e.g., oceanic plateau formation, accretion of juvenile mantle terrains to the continents, and basaltic underplating). How these accreted terrains were transformed into continental crust and how these retain geophysical properties of crustal material represent another central question that has been open to debate (the 'Ontong Java paradox' outlined later in the text). Crustal formation is related to heat generation in the Earth, which has clearly declined since early Archean time through the Proterozoic and Phanerozoic. What was the reflection of this change on the modes and mechanism of crust production, for example, intracrustal or lithospheric melting versus subduction-related magmatism and arc accretion?

In this chapter, we summarize some of the main results that emerge from the extensive efforts that were devoted over the past decades to determine the composition of the upper crust and lower crust and bulk continental crust (this subject and the relevant literature were thoroughly reviewed by Taylor and McLennan (1985, 1995, 1996) and more recently by Rudnick and Gao (2003)). Then, we go through the 'Armstrong–Moorbath debate' on early versus episodic crustal growth and the 'Reymer and Schubert dilemma' concerning arc–plume modes of crustal growth. We discuss mantle overturn and accretion models (cf. Ben-Avraham et al., 1981; Stein and Hofmann, 1994) that invoke episodic activity of large mantle plumes and accretion of juvenile mantle material to the continents and alternative arc-tectonics models (Patchett and Chase, 2002) or recent suggestions for continental preservation models (Condie et al., 2011; Hawkesworth et al., 2010). We discuss models of calc-alkaline and alkaline granitoids (the most important magma type comprising the upper continental crust).

Finally, we describe the production of juvenile continental crust in major segments of the continents (the 'major orogens'): the Archean Baltic Shield and Superior province, the Proterozoic Birimian orogen, and the late Proterozoic Arabian–Nubian Shield (ANS).

9.07.2 Structure and Chemical Composition of the Continental Crust

9.07.2.1 General

Based on seismic wave velocities, the continental crust can be vertically divided into two parts: the upper crust between the surface discontinuity and the Conrad discontinuity, which is poorly defined at 10–20 km from the surface, and the lower crust that principally extends from the Conrad discontinuity to the Moho discontinuity.

The problem in defining the location of the intracontinental crustal boundaries and obtaining an estimate of the chemical composition of the bulk continental crust reflects the complex history of the crust in different regions of the Earth. This evaluation requires knowledge on the type of the rocks comprising the upper crust and lower crust in various crustal domains. However, only a little portion of the continental crust is exposed (e.g., due to tectonic exhumation of

mainly the upper crustal segments and deep erosion) and accessible for direct research. The composition and physical properties of the lower crust are mainly evaluated from geophysical data combined with information derived from xenoliths and outcrops of uplifted granulite terrains.

Nevertheless, many researchers devoted extensive efforts over the years to base an estimate of the bulk chemical composition of the crust. This was done by determining the distribution and chemical composition of the dominant rock types that are exposed at the surface (e.g., the pioneering studies of Clarke and his coworkers Clarke, 1889; Clarke and Washington, 1924) and recent works by Gao et al. (1998) (see the comprehensive summaries of this topic by Condie (1993) and Rudnick and Gao (2003)). Another approach was to use sediments as natural samplers of large areas of the upper continental crust. Goldschmidt (1933, 1958) introduced this approach in his pioneering studies on glacial sediments from the Baltic Shield.

Later, other researchers applied similar ideas and methodologies using other samplers such as fine-grained sediments such as desert dust (e.g., loess) or deep-sea terrigenous sediments. Important examples are the works of Taylor and McLennan (1985), Plank and Langmuir (1998), Gallet et al. (1998), Barth et al. (2000), and Hattori et al. (2003). The results of these and other related studies were used to estimate the chemical composition of the upper crust (e.g., Rudnick and Gao, 2003; Taylor and McLennan, 1995).

The estimates on the bulk composition of the continental crust require, however, knowledge on the composition of the deeper parts of the crust. This information is provided mainly by xenoliths of mafic granulites that are transported to the surface by intraplate alkali basalts (mostly in rift-related environments) and tectonically uplifted blocks that consist of granulite facies metamorphic terrains. The post-Archean metamorphic granulite terrains appear to represent the upper crustal segments or transitional zones between the upper crust and lower crust (e.g., depressed in Himalayan-type collision zones (Bohlen and Mezger, 1989; Mezger, 1992)). Xenoliths of mafic granulites are considered as representatives of the lower crust. The xenoliths were interpreted as cumulates from basaltic magmas that were rising to lower crustal levels and were later subjected to the granulite facies metamorphism (cf. Kay and Kay, 1981; Rudnick, 1992). Since the xenolith data are rather sporadic, Rudnick and Fountain (1995) combined it with seismic velocities to get an estimate of the bulk lower crust composition. Nevertheless, it should be noted that the xenoliths of mafic granulite were mostly transported to the surface by rift-related alkali basalts and they may be associated only with the particular tectonic–magmatic environment that also supports alkali basalt production. Rifting environments are associated with the rise of asthenospheric (mid-ocean ridge basalt (MORB)-like) mantle into the spreading lithosphere (see the example of the opening of the Red Sea later in the text), and thus, the production of the source lithologies of the mafic granulite xenoliths may be associated with additions of new material to an older continental lithosphere. This scenario is supported by younger Re–Os isotope ages of lower crustal xenoliths compared to the above-lying lithosphere (Rudnick, 1992).

Rudnick and Gao (2003) combined the estimates of the upper crust, middle crust, and lower crust composition to produce a bulk continental crust composition. Trace element patterns (normalized to primitive mantle values) of the lower crust, upper crust, and bulk crust are illustrated in **Figure 1**.

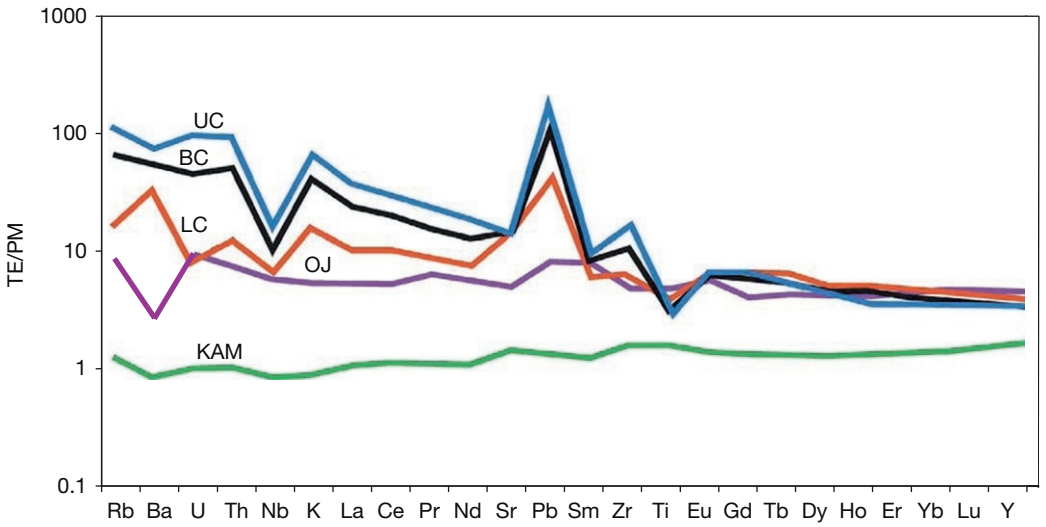


Figure 1 Primitive mantle-normalized trace element patterns of the upper, lower, and bulk crust estimates (UC, LC, and BC, respectively); the Cretaceous Ontong Java Plateau (OJ) basalt; and Archean Kambalda komatiite (KAM). Note the overall enrichment of the UC and BC crustal samples in the incompatible trace elements, the distinctive positive Pb and negative Nb anomalies in the crustal samples, and the flat patterns of the plateau and (almost primitive in KAM). A significant differentiation is required to transform OJ or KAM-type oceanic crust to the upper continental crust. The fractionation of trace elements like Nb and Pb could be furnished by processes in the subduction environment. Nevertheless, if accreted plateaus similar to OJ basalts comprise an important part of the continental crust, they either are not well represented by the bulk continental crust composition estimates or underwent (before or during accretion) internal magmatic differentiation (which are reflected by their continental-like seismic velocities; see **Figure 6**). Data sources: Rudnick and Gao (2003) and Mahoney et al. (1993a,b).

The order of appearance of the trace elements on the abscissa reflects the tendency of the incompatible trace element to move preferentially from the source to the melt during magma generation. This behavior can be applied to the processes that are involved in the production of continental crust from mantle sources (Hofmann, 1988). Noticeable features in this figure are the strong positive Pb anomalies and the negative Nb anomalies in the lower, upper, and bulk continental crust curves. These anomalies are probably related to mantle–crust differentiation processes that can fractionate Pb and Nb from the REE, Th, and U (cf. Hofmann et al., 1986). Such fractionation could occur in the subduction environment ('the subduction factory') along with the production of 'arc-type' calc-alkaline magmas (cf. Stein et al., 1997). The Mesozoic Ontong Java Plateau basalts and the Archean Kambalda komatiite display distinctly trace element patterns that are distinctly different from those of the 'lower, upper, and bulk' continental crust. These type of magmas show almost 'flat' primitive trace element patterns (although the Ontong Java sample displays small negative Pb and Ba anomalies that seem to oppose the strong positive Pb and Ba of the 'average lower crust'). If the Kambalda and Ontong Java type magmas represent important constituents of continental crust, they are not well represented by the 'bulk crust' patterns.

9.07.2.2 Composition of the Continental Crust: Open Questions

In the succeeding text, we outlined several topics and open questions that emerged from the extensive efforts to evaluate of the chemical composition of the continental crust.

9.07.2.2.1 Problems with the simple 'andesitic model'

The overall geochemical similarity of continental crust to arc-type magmas (e.g., the estimated average ~60% SiO_2 andesitic composition of bulk continental crust and similar values of indicative trace element ratios such as Ce/Pb and Nb/Th in arc–crust magmas) led many workers to suggest that the arc-subduction environment where andesitic magmas are produced is the main locus of juvenile crust production (the 'andesitic model' of continental crustal formation, e.g., Taylor, 1967) and that accretion of arcs at convergent margins is a major means of continental crustal growth. Transport and differentiation of melts and fluids that eventually control the composition of continental crust were associated with the subduction environment where mantle-wedge magmas and associated fluids with trace elements move upward while interacting and mixing with the bulk peridotitic assemblages (e.g., Kelemen, 1995). However, most arcs are too mafic in composition to build continental crust by simple accretion (cf. Anderson, 1982; Nye and Reid, 1986; Pearcy et al., 1990). Moreover, Taylor and McLennan (1985) pointed out several problems with a simplistic application of the andesitic model, such as the failure of this model to explain the abundances of Mg, Ni, and Cr and the Th/U ratio in upper crustal rocks. Rudnick (1995) proposed that the lower Sr/Nd and La/Nb ratios in continental crust compared to arc magmas are compatible with the involvement of intraplate magmatism in continental crust production. The andesitic magmatism and in a broader sense the intermediate 'granodioritic' bulk composition of the upper crust can be considered as a product of the differentiation processes that are involved in mantle–continental crust transformation. Taylor and McLennan (1985) suggested that only ~20–25% of the crust reflects the currently operating 'andesitic growth mode,' whereas the dominant mechanism that produced the Archean crust was the derivation of the bimodal basic–felsic magma suites from mantle sources (e.g., the formation of Archean tonalite–trondhjemite–granodiorite (TTG) magmas). It is argued later in the text that the production of thick oceanic crust and lithosphere by mantle plume activity and their accretion to the continents were an essential part of continental crust production throughout the past 2.7–2.9 Ga of the Earth history.

9.07.2.2.2 Plumes and arcs

The contribution of juvenile material from 'plume' or 'arc' sources to the continents can be assessed by using trace element ratios of Ce/Pb and Nb/Th that are distinctly different in arcs and plume-related mantle domains such as ocean island basalt (OIB) and oceanic plateaus (Hofmann et al., 1986; Kerr, 2013; Stein and Hofmann, 1994). Metatholeiites and greenstone sequences from major ancient orogens (e.g., the Superior, Birimian, and ANS) and plateau basalts such as the Ontong Java tholeiites display flat (chondrite-like) REE patterns (**Figure 2**) and where data exist show 'plume'-type Nb/Th ratios (**Figures 1** and **3**). It appears that these juvenile crustal units are not well represented by the sediment samplers. The upper crustal lithology and composition are largely dictated by the subduction and postaccretionary processes such as crustal melting and granite production, lithospheric delamination, and asthenospheric rising. These processes involved the preferential recycling of major elements (Mg and Ca) to the mantle at subduction zones and the removal of other major elements such as Si, Al, Na, and K to the continental crust (cf. Albarède, 1998). The subduction factory can be envisage as a big chromatographic column that fractionates the incompatible trace elements Nb, Pb, and Rb from REE and Th and U (Stein et al., 1997). Overall, these processes modified the

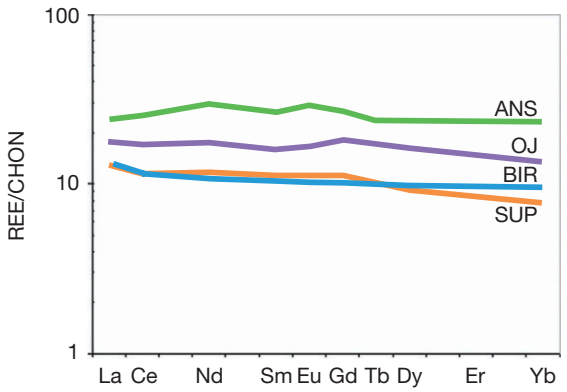


Figure 2 Chondrite normalized REE pattern in oceanic basalt from the Ontong Java Plateau (OJ) and several metatholeiites that were interpreted as plateau basalts from the Arabian–Nubian Shield (ANS), Superior province (SUP), and Birimian orogen (BIR). All these basalts were erupted rapidly and formed thick magmatic sequences, marking the early stage in a major orogenic cycle and production within oceanic basin. Data sources: Mahoney et al. (1993a,b), Abouchami et al. (1990), Reischmann et al. (1983), and Arth and Hanson (1975).

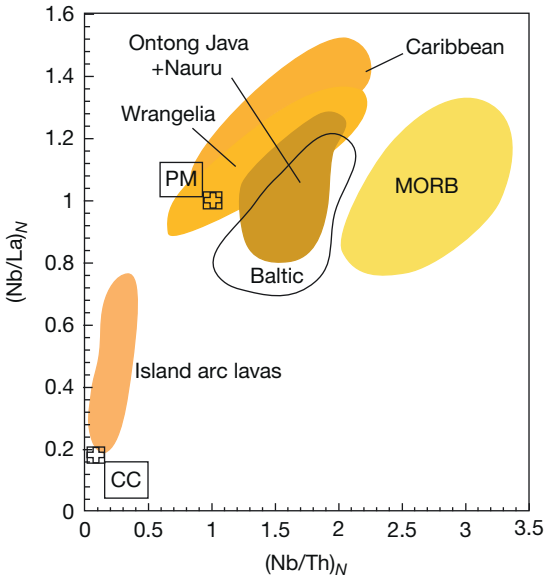


Figure 3 The location of basalts from several plateau basalt provinces on a $(Nb/La)_N$ versus $(Nb/Th)_N$ diagram (reproduced from Puchtel IS, Arndt NT, Hofmann AW, et al. (1998) Petrology of mafic lavas within the Onega plateau, central Karelia: Evidence for 2.0 Ga plume-related continental crustal growth in the Baltic Shield. *Contributions to Mineralogy and Petrology* 130: 134–53; Puchtel IS, Hofmann AW, Mezger K, Jochum KP, Shchipansky AA, and Samsonov AV (1998) Oceanic plateau model for continental crustal growth in the archaean, a case study from the Kostomuksha greenstone belt, NW Baltic Shield. *Earth and Planetary Science Letters* 155: 57–74). The Plateau basalts define fields that are distinctly different from island arc magmas and MORB and lie between primitive mantle (PM) compositions and Nb enriched that are similar to OIB and some rift-related alkali basalts (see **Figure 4**). Hofmann et al. (1986) attributed the high Nb/U and Nb/Th ratios to production of a 'residual' mantle after extraction of the low Nb/Th continental crust. Stein et al. (1997) proposed that Nb can be fractionated from REE, U, and Th due to production of amphibole 'front' in the mantle wedge, which they viewed as a giant chromatographic column. The Nb-rich zone is recycled back to the mantle or is frozen at the lower part of the lithospheric mantle after cessation of subduction. Could a similar process produce Nb/Th(REE) fractionation in the sources of the oceanic plateaus basalt? Does the convergence of several plateau basalts to primitive mantle ratios suggest that the plume head sample mantle with primitive Nb/Th–La ratios? In any event, Archean plateau magmas such as the Baltic Shield and the Phanerozoic magmas overlap to some extent in the diagram but are clearly not identical, suggesting some temporal changes in the Nb/Th ratios.

original geochemical signature of the juvenile mantle material and eventually shifted the upper continental crust to its average andesitic composition (Albarède, 1998; Rudnick, 1995; Stein and Goldstein, 1996).

9.07.2.2.3 The Eu dilemma

The application of REE to the study of the composition and evolution of the continental crust goes back to the works of Goldschmidt (1933), and the subject was thoroughly developed by Taylor and McLennan (1985). Most of the fine-grained sedimentary rocks that are used to estimate the bulk upper crust compositions typically show negative Eu anomaly (e.g., loess sample in **Figure 4**), pointing on significant

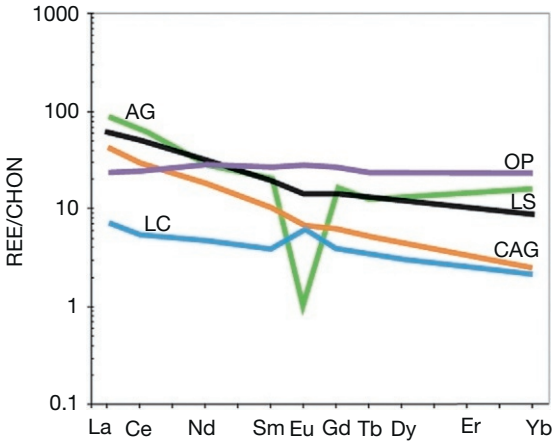


Figure 4 REE (chondrite normalized) pattern of magmas representing the main stages in the growth of a 'typical' orogen: tholeiitic magmas related to oceanic plateaus (OP), calc-alkaline granites (CAGs), and alkaline granites (AGs) as well as xenoliths of mafic granulites that are interpreted as lower crustal (LC) samples and late Pleistocene loess material that sample the exposed–eroded upper crust in the Sahara desert (LS). The pattern of the loess material is identical to bulk upper crust estimate (not shown). Note the appearance of a pronounced negative Eu anomaly related to plagioclase fractionation in the alkali granites and the almost smooth pattern of the CAGs. The production of the alkaline granites in the ANS was attributed to significant fractionation of a mafic assemblage that could be recycled to the mantle (Mushkin et al., 2003). The CAGs could be the product of midcrust (amphibolite-type) melting, which would produce residual lithology. The topic of production of vast calc-alkaline granitic batholiths is (surprisingly) not well understood. The production of granites requires water ('no water, no granites'), but the source of heat for the melting and the fate of the residual crust are not well known and the whole subject requires more studies.

intracrustal melting processes that require the existence of a residual reservoir with REE pattern showing positive Eu anomaly (Taylor and McLennan, 1995). However, such a reservoir remains largely unknown, because the xenoliths of mafic granulite that are considered as lower crustal samples and indeed show positive Eu anomalies are not interpreted as partial melting residues (cf. Rudnick., 1992; Rudnick and Taylor, 1987). It should be noted, however, that among the upper crustal rock suites, mainly the alkaline granites and related volcanics display distinctive negative Eu anomalies, while most of the calc-alkaline granites from juvenile crustal terrains display REE patterns with small or no Eu anomaly (e.g., the calc-alkaline batholiths of the Birimian and the ANS orogens; **Figure 4**). The alkaline granites and related rhyolitic volcanics typically characterize the closing stages of large orogenic cycles and their petrogenesis reflects fractionation of mafic magmas in shallow crustal levels (e.g., Mushkin et al., 2003, and see elaboration of this topic later in the text). These suites, however, are a rather small fraction of the total upper crust magmatic inventory that is probably irrelevant to the total crustal budget.

9.07.2.2.4 Crustal foundering

The search after 'residual reservoir' is associated with models of crustal foundering. This topic was discussed by Plank (2005) in light of the high Th/La ratio in the continental crust (e.g.,

Figure 1). She argued that the high Th/La ratio is unlikely to reflect fractionation processes within the subduction environment and neither could it be related to processes of intraplate magmatism or to Archean magmatism (e.g., Condie (1993) showed that Archean crust was rather characterized by low Th/La). Alternatively, the Th/La continental crust ratio can be related to the production of lower crustal cumulates (with low Th/La ratios) and foundering of such cumulates and partial melting restites (both represent high-density heavy crust) into the mantle. This process would create over geologic time bulk continental crust with high Th/La ratio.

9.07.2.2.5 Archean TTG magmatic suites

Archean magmatic suites were characterized by the bimodal mafic–felsic compositions with scarce appearance of andesitic rocks (Taylor and McLennan, 1985). Thus, the simple 'andesitic model' cannot be applied to the production of the major inventory of continental crust, which is Archean in age. The felsic members of the TTG suites show fractionated HREE patterns with no Eu anomaly, indicating melting in the garnet stability zone and no evidence for intracrustal melting involving plagioclase fractionation. The felsic precursors of these sediments were interpreted to reflect partial melting products of basaltic or TTG sources. The TTG for themselves were interpreted as fractionation products of intermediate parental magmas that were derived from enriched and metasomatized mantle peridotite (Shirey and Hanson, 1984) or melting products of the subducting slab (Moyen and Martin, 2012). For example, high-Al TTG plutons that are characterized by high La/Yb ratios, like those exposed in the Superior province greenstone–granitoid terrains, were interpreted as melting of young, hot subducting oceanic slabs (Drummond et al., 1996). In this relation, the production of modern magnesian andesites, Nb-rich basalts, and adakites could be attributed to the subduction of hot oceanic slabs (e.g., Central Andes; Stern and Kilian, 1996). The processes involved with the TTG production required an environment with high heat flow (supporting the production of Mg-rich basalts and komatiites) and probably rapid stirring of the mantle. Taylor and McLennan (1985) envisaged the formation of the Archean continental nuclei in a process involving eruption piling and sinking of tholeiitic basalts followed by partial melting and formation of the felsic magmas. To a limited extent, this scenario resembles the descriptions of fast production of thick oceanic crust in relation to mantle plume activity and the subsequent processing of the juvenile-enriched crust by subduction and intraplate processes possible since late Archean time. The major tectonic–magmatic element in this description that apparently was not significant during the Archean is the 'subduction crustal factory' and the production of andesitic–calc-alkaline magmas.

9.07.2.2.6 Post-Archean granites

While the felsic members of the Archean bimodal suites show highly fractionated HREE, the Post-Archean granitoids typically show flat or slightly fractionated HREE patterns, indicating a shallower depth of source melting above the garnet stability zone. The depth of melting thus represents a fundamental difference between Archean and post-Archean crustal formations (Taylor and McLennan, 1985). It should be stressed that TTG magmas are petrogenetically different from the post-Archean granitoids. For example, if subducting plate melts in the eclogite stability zone, there is a possibility to produce TTG magmas and thus granitic rocks without water. The post-Archean granitoids are melting and differentiation products of water-bearing mafic sources at shallower depth of the crust. This melting involves fractionation of amphibole instead of clinopyroxene and orthopyroxene. Since the amphibole is low in SiO_2, the residual melt is more silicic, driving the differentiation from basalt melt (the 'normal' melt produced in subduction zones) to silicic melt all the way to granitic melt. The apex of this discussion is the requirement of water-carrying subduction processes in the formation of post-Archean granitic magmas.

Campbell and Taylor (1983) emphasized the 'essential role of water in the formation of granites' and, in turn, continents and introduced to the literature the important concept of 'no water, no granites; no oceans, no continents.' They argued that the Earth, which is the only inner planet with abundant water, is the only planet with granite and continents, whereas the Moon and the other inner planets have little or no water and no granites or continents. This concept could be extended to 'no water, no plate tectonics, no continents,' since with no water, plate tectonics is unlikely to work. Thus, there is a fundamental linkage between the formation of granitic (granodioritic) continental crust and plate tectonic mechanism.

9.07.2.3 The Lithospheric Mantle and Continental Stability

Beneath the Moho and the continental crust lies the lithospheric mantle whose thickness can reach 300–400 km beneath old continents and is relatively thin (up to 100–120 km) beneath young continents (e.g., beneath the Canadian and ANS, respectively). Subcontinental lithospheric xenoliths show large variations and heterogeneities in the trace element abundances that over time can 'incubate' large variations in the radiogenic isotope systems such as Pb, Nd, and Sr (cf. Hawkesworth et al., 1983; McKenzie and O'Nions, 1983; Stein et al., 1997). The magmatic and thermal history of the lithospheric mantle is related to that of the above-lying continental crust. Based on Pb, Nd, and Sr isotopes in ophiolites, galenas, and basalts, Stein and Goldstein (1996) suggested that the lithospheric mantle beneath the Arabian continent was generated during the events of the late Proterozoic crustal formation. The lithospheric mantle can be regarded as a 'frozen' mantle wedge, which previously accommodated the production of calc-alkaline magmas and fractionation–transportation processes of trace elements on route to shallower levels of the crust (the subject is elaborated in the succeeding text). After cessation of subduction, the lithospheric mantle (which also contains water) could become a source for alkali magmas that rose to the shallower levels of the continental crust (e.g., alkali granites) or erupt at the surface (cf. Mushkin et al., 2003; Weinstein et al., 2006). Thus, the lithospheric mantle not only is a potential source of information on the 'subduction factory' processes but also can be regarded as an important source of magmas that constitute a part of the continental crust inventory. Yet the source of heat and conditions of lithospheric melting is not obvious. One possibility is that the asthenospheric mantle that rises into the rifting continent and the thinning lithosphere supplies

the heat for the melting. The production of magmas caused eventually the depletion of the lithospheric mantle from incompatible trace elements and basaltic components and eventually led to its exhaustion.

9.07.3 Models of Crustal Formation and Continental Growth

9.07.3.1 General

Models that describe the mechanism of continental crustal growth and the evolution of continental masses can be divided into two major groups: (1) models of early crustal growth and subsequent recycling and (2) models of continuous or episodic crustal growth.

The most prominent researcher that pushed the 'early crustal growth/recycling' model was Armstrong. He proposed that soon after the early mantle–crust differentiation and formation of continental crust, the Earth reached a steady state with accretion and destruction of continental masses occurring in approximately equal rate with essentially constant volumes of ocean and crust through geologic time (Armstrong, 1981a, 1991). The key process in this model is the recycling of the continental masses back into the mantle. Armstrong based his model on the principles of the constancy of continental freeboard and the uniformity of thickness of stable continental crust with age. These two critical parameters imply negligible continental crustal growth since 2.9 Ga BP. Defending his model, Armstrong addressed part of the arguments that were put against the steady-state model, mainly those that relied on the behavior of the radiogenic isotopes (e.g., Nd, Sr, and Pb; Armstrong, 1981b).

Among the pioneering researchers that developed the ideas of 'continuous/episodic crustal growth' were Rubey (who advocated the underplating theory as well), Engel, Hurley, and Taylor. The prominent researcher of this school of thinking was Moorbath who pointed on the difficulty to subduct large amounts of continental crust due to its buoyancy and argued that the chronological evidence derived from continental crustal rocks suggests the creation of juvenile continental crust throughout the Earth history. Moorbath also developed the concept of continental accretion as a major building process of continental crust (Moorbath, 1975, 1978).

The steady-state recycling theory of crustal growth requires evidence for substantial sediment recycling. Sediments are clearly recycled in the subduction environment. Various geochemical tracers of arc magmas demonstrated sediment involvement in subduction: for example, the radiogenic isotope ratios of Nd, Pb, Sr, and Hf (cf. Patchett et al., 1984; White et al., 1985), ^{10}Be isotopes (cf. Tera et al., 1986) and trace element fluxes (Plank and Langmuir, 1998). Plank (2005) introduced the Th/La ratio as a tracer of sediment recycling in subduction zones. Th/La in the subducting sediments reflects the mixing between terrigenous material with high 'upper crustal' Th/La ratio (0.3–0.4) and metalliferous sediments and volcaniclastic sediments with low Th/La ratios (<0.1). White and Patchett (1984) argued that a significant but small (at least 1–2%) contribution of older continental material could be potentially recycled into new continents. Nevertheless, based on mass–age distribution and Nd isotope data of sediments, McLennan (1988) estimated the mass of sediment available for subduction to be $<1.6\times10^{15}$ g year^{-1} and argued that that is insufficient to support a steady-state crustal mass according to the Armstrong model (see also Taylor and McLennan, 1995).

Hofmann et al. (1986) suggested that the Nb/U ratio in mafic–ultramafic magmas can be used to deduce the history of continental crustal growth because Nb/U ratio changed from the primitive mantle ratio of ~30 to a higher ratio of ~47 in the 'residual' mantle due to extraction of crust with Nb/U of ~10. Some of the Archean komatiites (e.g., samples from Kambalda and Kostomuksha provinces) show high Nb/U ratios of ~47 (Puchtel et al., 1998a,b; Sylvester et al., 1997) that may support the Armstrong steady-state model. Yet other komatiites and early Proterozoic tholeiites show Nb/U ratios that are similar to the primitive value (**Figure 1**). Puchtel et al. (1998a,b) commented on this topic that "the Nb/U data alone cannot unequivocally resolve this important controversy."

Although the continents appear to have undergone episodic changes throughout all of geologic time, the available geochronology of crustal formation shows time periods when large areas of continental crust were formed at a fast rate, alternating with apparently more quiescent periods of low crustal formation rates (marked in **Figure 5**). It may be that these quiescent periods merely represent missing continental material because the crust was destroyed and recycled into the mantle (according to the Armstrong model) but actual evidence for this process is indirect at best.

9.07.3.2 Timing and Rates of Crustal Growth

9.07.3.2.1 General

Crustal formation ages were determined by direct dating of major crustal terrains (numerous works applying the Rb–Sr, Sm/Nd, and U–Pb dating methods) and by using the concept of Nd – crustal model ages that mark the time of major chemical fractionation and the Sm/Nd ratio change accompanying the extraction of melts from the mantle and their incorporation into the continental crust (e.g., DePaolo, 1981; DePaolo et al., 1991; McCulloch and Wasserburg, 1978; Nelson and DePaolo, 1985).

The Sm/Nd crustal formation ages suggest that 35–60% of the currently exposed crustal masses were produced in the Archean (while actual exposed Archean crust accounts only for 14%; Goodwin, 1991). **Figure 5** shows that the most dramatic shift in the generation of continental crust happened at the end of the Archean, 2.7–2.5 billion years ago. The late Archean early Proterozoic pulse of crustal growth is represented in all continents (Condie, 1993). Before 3 Ga, very little continental crust had been preserved. The oldest preserved terrestrial rocks are the 3.96 Ga Acasta gneiss in the Northwest Territories of Canada (Bowring et al., 1990), and detrital zircons from the Yilgarn block, Western Australia, were dated to a much older age of 4.4 Ga (Wilde et al., 2001). The evidence from Hf and O isotopes in zircons on the oldest continental crust is discussed in the succeeding text.

9.07.3.2.2 Ages of the major orogens

The available geochronology of crustal formation shows time periods when large areas of the continental crust were formed

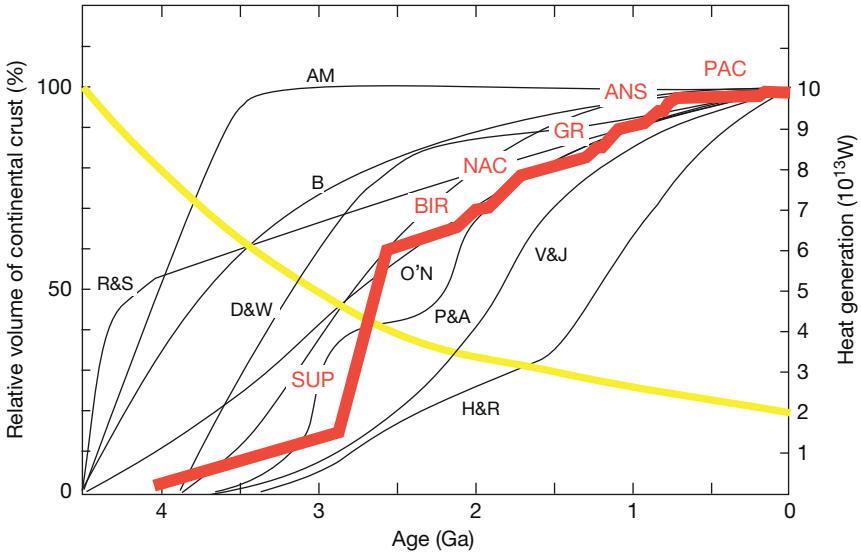


Figure 5 Models of continental crust (cumulative) growth and heat generation throughout the Earth history (reproduced from Reymer A and Schubert G (1984) Phanerozoic addition rates to the continental-crust and crustal growth. *Tectonics* 3: 63–77; Taylor SR and McLennan SM (1985) *The Continental Crust*: *Its Composition and its Evolution*. Oxford: Blackwell; Taylor SR and McLennan SM (1995) The geochemical evolution of the continental-crust. *Reviews of Geophysics* 33: 241–265; Condie, 1990; Patchett PJ (1996) Crustal growth – Scum of the Earth after all. *Nature* 382: 758–759). Crustal curves: R&S, Reymer and Schubert (1984); AM, Armstrong (1981a,b); B, Brown (1979); D&W, Dewey and Windley (1981); O'N, O'Nions et al. (1979); P&A, Patchett and Arndt (1986); V&J, Veizer and Jansen (1979); H&R, Hurley and Rand (1969). The thick red curve is after Taylor and McLennan (1995), with marks of the major orogenic episodes that were related by Stein and Hofmann (1994) to major upwelling events and plume head uprise in the mantle (the MOMO model illustrated in **Figure 9**). SUP, Superior province; BIR, Birimian orogeny; NAC, North Atlantic Continent; GR, Grenville orogeny; ANS,Arabian–Nubian Shield; PAC, Pacifica 'superplume.'

in a fast rate, alternating with apparently more quiescent periods of low crustal formation rates. The recent advances in U–Pb dating of zircons provided extensive sets of data that indicate the times of rapid growth of continental crust. For example, Condie and Aster (2010) and Condie et al. (2011) compiled numerous U–Pb ages of detrital zircons from all continents and whole rock granites and showed clustering of the ages at prominent peaks at 2.7, 2.5, 2.1, 1.9, 1.1, and 0.6 Ga. According to the Moorbath episodic growth model, these age peaks describe the timing of the major crustal formation events that are termed here as 'the major orogens' (**Figure 5**). Armstrong would interpret the quiescent periods between the intervals of the 'major orogens' as representing missing continental material because the crust has been destroyed and recycled into the mantle. Another view that emerged in the last decade relates the age peaks to the times of collision events of the continents and enhanced preservation of continental crust during the assembly of major continents (e.g., Hawkesworth and Kemp, 2006; Hawkesworth et al., 2009, 2010; Kemp et al., 2006). Yet, the detailed histories of the major orogenic cycles show that juvenile crust addition to the crust from the mantle occurs in the early stage of the cycle during the production of oceanic plateaus (which can accrete to the older continents) or during the time interval (e.g., >200 Ma) of subduction-related calc-alkaline magmatism. These processes are illustrated later in the text by the history of the late Proterozoic ANS, which is possibly the best-exposed and preserved juvenile crustal terrain on Earth and is described in some details later in the text. Moreover, it was argued by Stein and Goldstein (1996) that all major orogenic cycles such as the Superior province, the Birimian orogen, and the ANS evolved through similar histories that begin with oceanic plume-related magmatism and continued with the development of subduction margins and related calc-alkaline magmatism followed by anatexis of the crust and production of calc-alkaline and alkaline granites. Most of the zircon ages of the ANS orogeny mark the time of granitoid production by anatexis of the previously formed juvenile crust (calc-alkaline granites) or lithospheric mantle (alkaline granites).

9.07.3.3 The Reymer and Schubert Dilemma

Reymer and Schubert (1984, 1986) evaluated the rate of growth of major continental crustal segments (e.g., the Superior province in North America and ANS) and showed that they significantly exceed the rate of continental addition that prevailed along the subduction margins during the Phanerozoic time (~1 km^3 $year^{-1}$). The Phanerozoic rate was based on estimates of crustal additions along Mesozoic–Cenozoic arcs, hot spots, and some other additional sources (e.g., underplating). They calculated the total addition rate (mainly along the arcs) of 1.65 and total subtraction rate 0.59 km^3 $year^{-1}$, yielding a net growth rate of 1.06 km^3 $year^{-1}$. In addition, Reymer and Schubert calculated a growth rate by an independent model based on the constancy of freeboard relative to the mantle with declining radiogenic heat production (Schubert and Reymer, 1985).

The term freeboard was derived from civil engineering, describing the additional height above a normal operating water level and the top of the water-holding structure. The geologic–geophysical use of this concept encompasses the complicated relation between continental crustal thickness,

volume, mantle temperature, and the Earth's heat budget. Wise (1974) tied the freeboard concept to crustal growth rates, by suggesting that approximately constant continental crustal volumes and areas pertained since the Archean–Proterozoic boundary, thus maintaining approximately the same elevation or freeboard of the continents above mean sea level. The freeboard calculation of Reymer and Schubert yielded a Proterozoic–Phanerozoic growth rate of 0.9 $km^3\ year^{-1}$, very similar to the calculated value in the preceding text. The Phanerozoic growth rate is three times less than the average Archean growth rate. Assuming formation between 900 and 600 Ma during the Pan-African orogeny, Reymer and Schubert showed that if the ANS was generated by convergent margin magmatism, it would require an addition rate of 310 $km^3\ km^{-1}\ Ma^{-1}$ of convergence, compared with a global total of ~40 $km^3\ km^{-1}\ Ma^{-1}$ (Figure 6). At the very least, the results of Reymer and Schubert indicate that the local crustal formation rates of the major continental shields have as high as or higher than the present-day rate averaged over the entire globe. As a consequence of their analysis, Reymer and Schubert proposed that juvenile additions to continental crust require additional mechanism besides the arc magmatism and this mechanism could involve activity of hot spots or mantle plumes. The geodynamic conditions during these intervals of juvenile crust creation may resemble to some extent the Archean conditions.

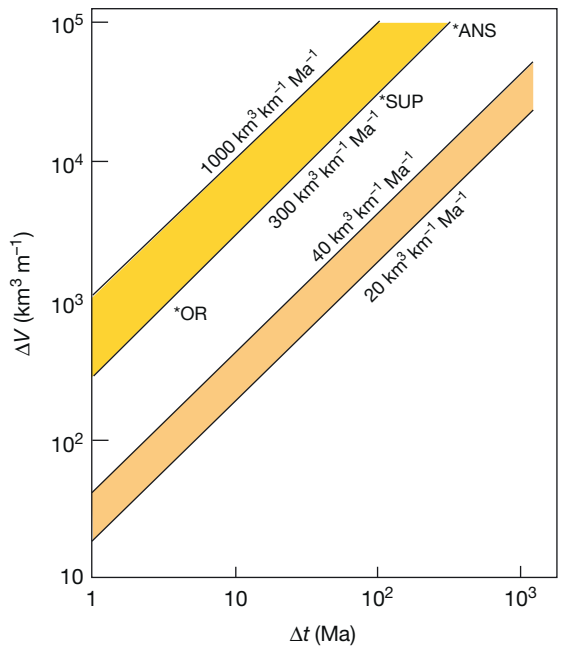


Figure 6 Arc crust (volume) addition curves (reproduced from Reymer A and Schubert G (1984) Phanerozoic addition rates to the continental-crust and crustal growth. *Tectonics* 3: 63–77) showing the large difference between Phanerozoic arc activity (20–40 $km^3\ km^{-1}\ Ma^{-1}$) and the growth of some major orogens (e.g., SUP (Superior province), ANS (Arabian–Nubian Shield) ~300 $km^3 km^{-1}\ Ma^{-1}$ OR (Oregon Coastal Range)). The figures demonstrate the 'Reymer and Schubert dilemma' that points on the necessity to invoke additional sources and mechanism to support the production of continental crust during the episodes of major orogen production.

Galer (1991) performed another evaluation of the interrelationships between continental freeboard, tectonics, and mantle temperature. He concluded that before ~3.8 Ga, the potential temperature exceeded 1600 °C in the shallow mantle and therefore little or no hypsometric or tectonic distinction existed between 'continental' and 'oceanic' regimes. Thus, the pre-3.8 Ga period was subjected to tectonic and sedimentary processes that were distinctly different to those characterized by the Earth surface since mid-Archean time. An important component in his evaluation is the assumption that the early Archean oceanic crust was significantly thicker than typical present crust. The Reymer and Schubert (1984, 1986) assessment indicates that mantle plumes could play an important role in post-Archean crustal history of the Earth. Later in the text, we discuss the possible role of thick oceanic crust (e.g., the oceanic plateaus) in continental crustal growth. We can ask the question whether post-Archean episodes of crustal growth have some important components that 'mimic' to some extent the early Archean conditions.

9.07.3.4 Continental Growth and Heat Production

The heat production in the Earth, which is the main driving force for mantle convection and upwelling processes, has declined by a factor of 5 since the early Archean time (see yellow curve in Figure 5). This decrease reflects the decay of the long-living radiogenic isotopes. Figure 5 indicates that continental crust accumulation does not reflect the Earth thermal history very well particularly in its earliest Archean history only since the late Archean a rough correspondence appears between the heat generation and crustal growth (Patchett, 1996). The heat engine of the mantle should have supported extensive mantle movements such as plume head upwelling in the Archean. Yet the low cumulative crustal growth in the Archean may indicate that the system was as efficient at destroying continental crust as making it (Taylor and McLennan, 1995).

The intensive pulses of continental crustal formation in the late Archean and early Proterozoic probably took place in a mantle that convected less vigorously than at earlier times. One possibility is that the tectonic regime became more similar to modern plate tectonics, where water played an essential role. By late Archean and early Proterozoic times, we might enter the regime of 'water, plate tectonics, continents.'

9.07.4 Oceanic Plateau, Accreted Terrains and Juvenile Crust Additions

9.07.4.1 General

With the advances made in geophysical techniques over the last 50 years, a different and new picture of the Earth began to emerge. For the first time, it was possible to image and sample the planet's crust and even layers deeper down. During the 1970s, it started to become clear that the ocean's floor is not homogenous, but contains areas of significantly thicker crust than the norm. The discovery of one such area, the Ontong Java Plateau with a crustal thickness of over 30 km, gave rise to the term 'oceanic plateau' (Kroenke, 1974). Today, about 100 such anomalous regions ranging in size from 1000 km to a few kilometers have been documented (Figure 7).

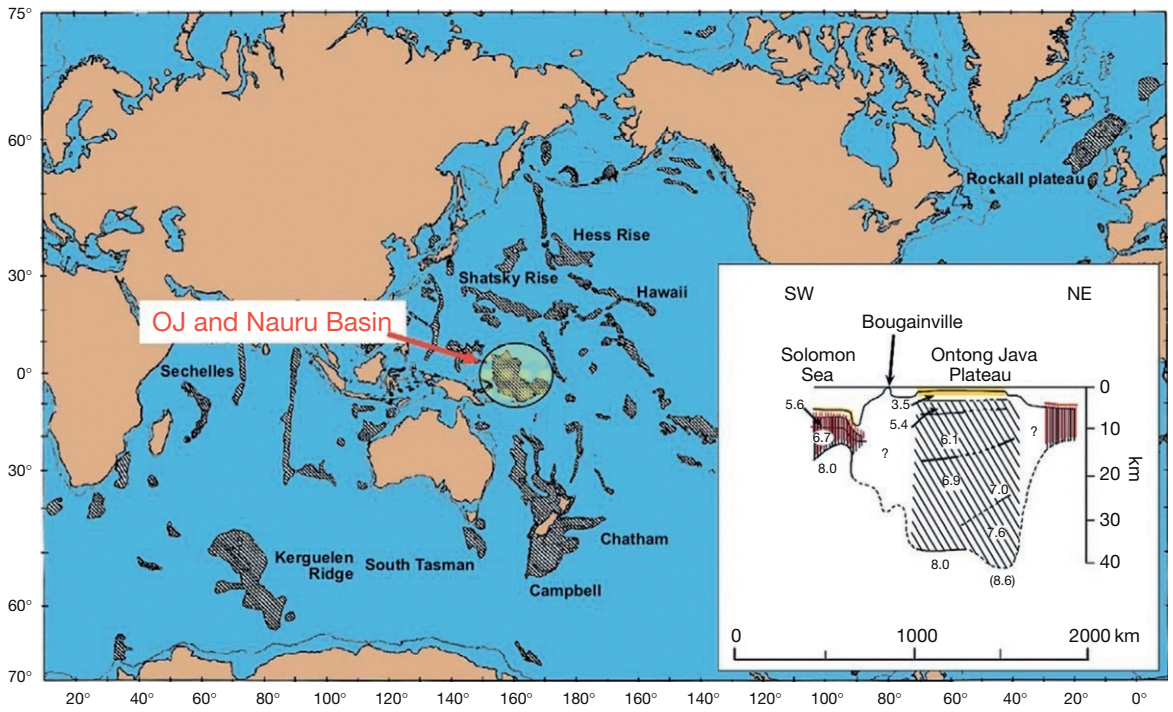


Figure 7 Distribution of oceanic plateaus and location of the Ontong Java Plateau and Nauru Basin in the Pacific. The Mesozoic Pacific was the locus of enhanced magmatism that was related to plume activity ('The latest pulse of Earth,' Larson, 1991). It also appears that shortly after the production of the thick oceanic plateaus, subduction zones were developed on their margins. The Nauru Basin comprised oceanic crust that erupted in ocean spreading environment but show geochemical similarities to the adjacent Ontong Java Plateau. Thus, it appears that both regions derived their magmas from similar enriched sources. This configuration supports the existence of an enriched upper mantle beneath the Pacific. Stein and Hofmann (1994) argued that the Nd–Sr isotopic composition of this mantle is similar to the PREMA isotopic composition (e.g., ϵNd ~+6 and $^{87}Sr/^{86}Sr$ ~0.7035 as defined by Wörner et al. (1986)) for various volcanic rocks around the world. It appears that the Pacific–PREMA composition characterized the entire upper mantle during the time of the production of the Ontong Java Plateau and Nauru Basin basalts and thus could mark the overturn or plume uprising events according to the MOMO model. Anderson (1994) proposed an alternative interpretation to the production of the Mesozoic Pacific plateau that requires no plume activity. He suggested that the upper mantle of the Mesozoic Pacific and the Indian Ocean is hot because this region has not been cooled by subduction for more than 200 Ma. Anderson argued that the timing of formation and location of the plateaus are related to the organization of the tectonic plates. This would require that the PREMA-type enriched isotopic compositions characterized the upper mantle rather than MORB, but magmatism in evolved rifting environments (e.g., Ross Sea–Antarctica and Red Sea–Arabia) clearly shows the transition from enriched isotopic composition to typical depleted MORB (e.g., Rocholl et al., 1995; Stein and Hofmann, 1992, and see **Figure 13** and 'Red Sea model' in **Figure 14**). (b) Seismic velocities across the Ontong Java Plateau. Surprisingly, the velocities suggest 'continental crust' composition and structure. This observation illustrates a kind of 'paradox' since it is expected that Ontong Java Plateau will show seismic structure of a thick oceanic crust. Does it mean that 'crustal formation' processes operate in the mid-oceanic environment? We regard the 'Ontong Java paradox' as a major open question that should be addressed by future studies.

The formation of these plateaus is still not clearly understood. They were considered to represent remnants of extinct arcs, abandoned spreading ridges, detached and submerged continental fragments, anomalous volcanic piles, mantle plume head traces, and uplifted oceanic crust. However, both isotope dating and isotope modeling support the theory that they formed rapidly, often in <2–3 My (Richards et al., 1989), with the majority forming at or near mid-ocean ridges (Kerr, 2013). These areas are conductive to large decompression melting (Eldholm and Coffin, 2000), providing a source of magma needed to form these features.

At least from a geophysical point of view, values of crustal seismic velocities, V_P, in the upper 5–15 km (where V_P ranges from 6.0 to 6.3 km s^{-1}; insert **Figure 7**), together with the anomalous thickness, relief, and gravity data, would indicate a continental structure for some of these plateaus (Ben-Avraham et al., 1981; Carlson et al., 1980; Nur and Ben-Avraham, 1977, 1978). Other plateaus are from clear oceanic origin, probably the result of continuous extrusion of basalts, probably from an active hotspot or mantle plume head.

Nur and Ben-Avraham (1977) and Ben-Avraham et al. (1981) suggested that the plateaus with the 'crustal seismic velocities' (e.g., Ontong Java Plateau) represent continental fragments rifted away from a parent continent that they named Pacifica. Stein and Hofmann (1994) and Abbott and Mooney (1995) on the other hand suggested that the plateaus were first extracted from the mantle within the oceanic basins above mantle plumes and were later accreted to the continents. In this context, it is interesting to note the coappearance of the basaltic sequences of the Ontong Java Plateau and those of the adjacent Nauru Basin in the Pacific Ocean (cf. Mahoney et al., 1993a,b). The tectonic setting of the Nauru Basin is clearly of a mid-oceanic spreading center type, with the typical appearance of magnetic anomalies and eruption of tholeiitic basalts.

The important point is that the Ontong Java Plateau and Nauru Basin basalts display similar geochemical characteristics such as Nd and Sr isotope ratios that are enriched relatively to the depleted MORB-type mantle (e.g., the basalts from the East Pacific Rise). Thus, it appears that the basalts erupted in the Nauru Basin and the Ontong Java Plateau represent a similar uppermost mantle that is enriched relative to the depleted (MORB-type) asthenospheric mantle. This enriched asthenosphere could be related to a large plume head (or a mantle upwelling) event that brought enriched material to the shallow mantle.

Richardson et al. (2000) produced a three-dimensional tomographic model of the seismic structure beneath the Ontong Java Plateau that indicated the existence of a low-velocity mantle 'root' reaching the depth of ~300 km. This was interpreted as a remnant of the Cretaceous Ontong Java plume that was attached to the plateau. Ishikawa et al. (2004) addressed the question of this seismically anomalous low-velocity root beneath the Ontong Java Plateau and its lithospheric mantle composition and structure by studying suite mantle xenoliths (peridotites and pyroxenites) from Malaita, Solomon Islands. The shallower mantle (Moho to 95 km) is composed of variably metasomatized peridotite with subordinate pyroxenite derived from metacumulates, while the deeper mantle (95–120 km) is represented by pyroxenite and variably depleted peridotites. The shallower and deeper zones are separated by a garnet-poor zone (90–100 km), which is dominated by refractory spinel harzburgites. Ishikawa et al. attributed this depth-related variation to different degree of melting for a basalt–peridotite hybrid source at different level of arrival depth within a single adiabatically ascending mantle plume: the lack of pyroxenites at shallower depths was related to extraction of hybrid melt from completely molten basalt through the partially molten ambient peridotite, which caused the voluminous eruption of the Ontong Java Plateau basalts. The authors concluded that the lithosphere forms a genetically unrelated two-layered structure, comprising shallower oceanic lithosphere and deeper impinged plume material, which involved a recycled basaltic component, now present as a pyroxenitic heterogeneity. The possible relevancy of this model to the shallower seismic structure of the Ontong Java Plateau (**Figure 7**) could be that evolved magmas were produced within the shallower layers.

Regardless of the mechanisms responsible for the formation of oceanic plateaus, they are assumed by many researchers to have an important role in the formation of new continental crust. Because they are part of moving plates and are embedded in them, they are destined to arrive at subducting plate boundaries and to accumulate, at least in some cases, onto the continental margin as accreted allochthonous terrains.

The potential of oceanic crust to be accreted to the continents rather than being subducted depends on its magmatic and thermal history and is reflected in its buoyancy. During 'normal seafloor spreading operation,' the oceanic lithosphere begins its history as a buoyant hot plate that cools with time and becomes susceptible for subduction. The buoyancy of the oceanic lithosphere depends on its age and density distribution (Oxburgh and Parmantier, 1977). The density distribution reflects the composition and the thickness of the crustal and mantle layers of the lithosphere, which in turn are controlled by the mantle temperature at place of crustal formation. Hotter mantle will produce thicker oceanic crust with more depleted lithospheric mantle, which can become less dense than more fertile mantle and, thus, more buoyant (Langmuir et al., 1992; McKenzie and Bickle, 1988; Oxburgh and Parmantier, 1977). This is probably the case in the production of oceanic plateaus by rising plume heads. The plumes are ~200–300 °C hotter than ambient mantle and can produce 15–40 km thick oceanic plateaus (compared with ~7 km thickness of normal oceanic crust). Therefore, some oceanic plateaus are too buoyant to subduct, and they can either be obducted into the continent (e.g., Caribbean and Wrangelia; Kerr et al., 1997; Lassiter et al., 1995) or, when arriving to the subducting margins, start subduction in the opposite direction (e.g., Ontong Java Plateau; Neal et al., 1997). The Archean mantle whose temperature was a few hundred degrees higher than the post-Archean mantle could produce thicker oceanic crust of 20–25 km (Sleep and Windley, 1982) and the plume-related plateaus could reach an average thickness of ~30 km (e.g., Puchtel et al., 1997). Some of these thick oceanic plateaus might be too warm and too buoyant when they reach the subduction margins and remain susceptible to subduction. Cloos (1993) and Abbott and Mooney (1995) argued that oceanic plates with crust thickness over ~25 km are unsubductable regardless of their age.

Nevertheless, several oceanic plateaus are currently being consumed, either by collision or by subduction. Among these are the Nazca and Juan Fernández Ridges off South America and the Louisville and Marcus–Necker Rises in the western Pacific (Cross and Pilger, 1978; Nur and Ben-Avraham, 1981; Pilger, 1978). These plateaus, which are presently being subducted, can clearly be associated with gaps in seismicity of the downthrown slab (e.g., Nur and Ben-Avraham, 1982). However, even more pronounced than these gaps are gaps in volcanism, associated with these plateaus. Close spatial association was found between the zones of collision or subduction of plateaus and gaps in volcanism in the Pacific (McGeary et al., 1985). Thus, these currently active areas of continental accretion are clearly identifiable.

The interaction of oceanic plateaus and other allochthonous terrains (e.g., island arcs, submarine ridges, and continental fragments) with convergent margins was recently evaluated by numerical models and experiments. For example, Mason et al. (2010) demonstrated that oceanic plateaus interacting with the trench of a subduction zone may alter trench behavior significantly. The plateau pins the trench, irrespective of slab and plateau rheology. Tetreault and Buiter (2012), using geodynamic numerical experiments, demonstrated how crustal properties of the allochthonous terrains impact the amount of the crust that is accreted or subducted, the type of accretionary process, and the style of deformation on the overriding plate.

9.07.4.2 Accretion of Allochthonous Terrains at the Pacific Margins

Large portions of the Pacific margins, especially the northeast margin, are made of accreted or allochthonous terrains (**Figure 8(a)**). The best understood accreted terrains are those in the northern cordillera of western North America, particularly in southern Alaska and British Columbia (Coney et al.,

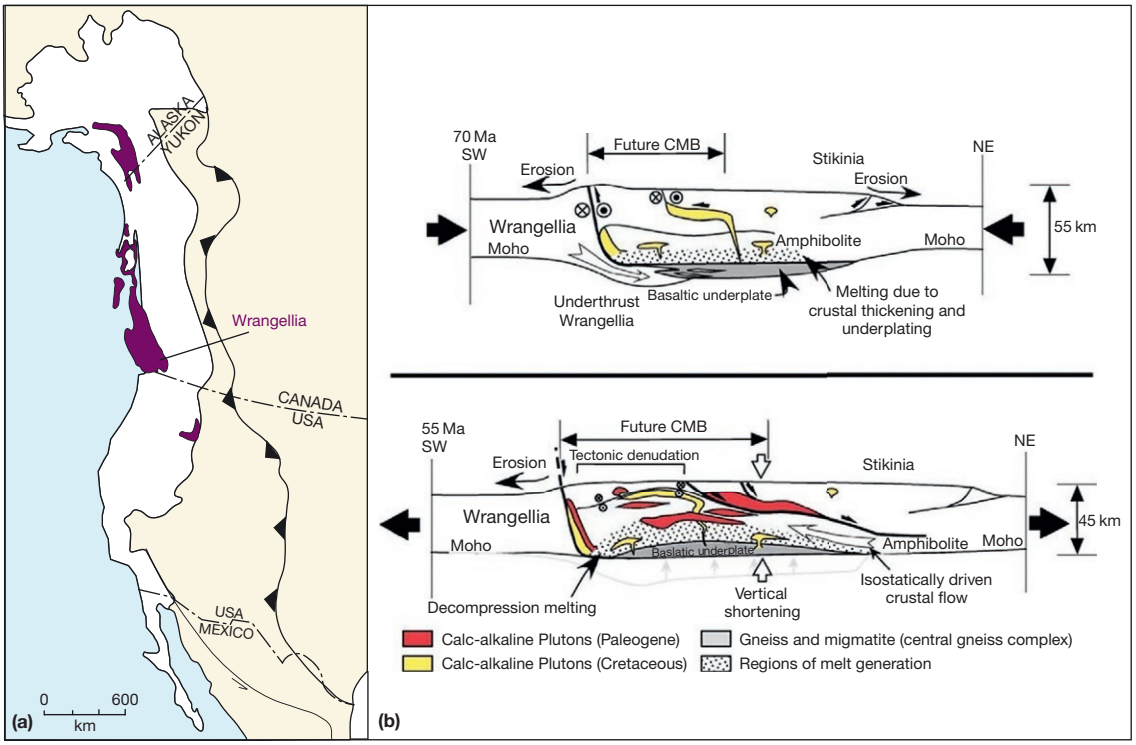


Figure 8 (a) Accretion plateaus along the western margins of North America and Alaska. The location of the Wrangellia oceanic plateau is marked in brown (reproduced from Ben-Avraham Z, Nur A, Jones D, and Cox A (1981) Continental accretion and orogeny: From oceanic plateaus to allochthonous terrains. *Science* 213: 47–54). (b) Model of plateau accretion and transformation of the oceanic crust to new continental crust and lithosphere (reproduced from Hollister and Andronicos, 2006). The figure shows processes of accretion, shortening and melting. Note the stratigraphic organization of Cretaceous and the Paleogene CAG and location of zones of melting due to crustal thickening and basaltic underplating. Amphibole appears above the Moho and its production could control Nb/Th(REE) fractionation (Stein et al., 1997).

1980; Monger, 1993). Paleomagnetic evidence indicates that many of the North Pacific accreted terrains in Alaska and northeast Asia migrated several thousand kilometers over periods of tens of millions of years prior to their accretion to the margins. The accreted terrains comprising the Canadian–Alaskan Cordillera are predominantly juvenile in composition (e.g., Patchett and Samson (2003) estimated that ~50% of the mass of the Canadian segment of the cordillera was juvenile crustal material). This evaluation is based on the initial $^{87}Sr/^{86}Sr$ and ϵNdt values of lithologic assemblages from both outboard and inboard terrains (e.g., Alexander, Cache Creek, and Slide Mountain as well as arc assemblages like Quesnel; Armstrong, 1988; Armstrong and Ward, 1993; Samson et al., 1989; 1990, 1991; and other references listed by Patchett and Samson (2003) in their recent review on this topic).

The Cache Creek terrane (CCT) of the Canadian Cordillera has an interesting history. It consists of accreted seamounts that originated adjacent to the Tethys Ocean in the Permian (Johnston and Borel, 2007). The paleogeographic constraints provided by the CCT indicate that much of the Canadian Cordilleran accretionary orogen is exotic. The accreting crustal block, a composite ribbon continent, grew through repeated collisional events within Panthalassa prior to docking with the North American Plate. Johnston and Borel (2007) suggested that Middle to Upper Triassic accretion of the Tethyan Cache Creek seamounts to Stikinia–Quesnellia and Lower Jurassic collision of Stikinia–Quesnellia with the combined pericratonic belt–Cassiar platform occurred >10 000 and >4000 km west of autochthonous North America, respectively. Most of the Cordilleran orogen is, therefore, exotic with respect to North America. According to their analysis, the development of the Cordilleran accretionary orogen was not, therefore, the result of sequential terrain accretion along the continental margin but was instead a two-stage process and resulted in a significant growth of the North American continent. Stage one involved the 230–150 Ma accretionary amalgamation of a ribbon continent within central to western Panthalassa. Stage two, from 150 to 55 Ma, consisted of collision, margin-parallel translation, and final coalescence of the composite ribbon continent with North America. The evolution and development of the Canadian Cordillera imply that accretionary orogens may, therefore, involve an offshore amalgamation stage that may take place far removed from the autochthonous continent.

In a detailed analysis of northern Cordilleran terrains, Colpron et al. (2007) suggested that although Cordilleran terrains are defined on the basis of their distinct tectonostratigraphic records, they are not necessarily limited to fault-bounded crustal blocks. They show that adjacent terrains can locally share depositional and geodynamic ties and that the nature of terrain boundaries evolves through time. A primary result of their analysis is that the Intermontane terrains represent one interrelated set of arcs, marginal seas, and continental

fragments that once formed a Paleozoic to early Mesozoic fringe to North America, the peri-Laurentian realm. By contrast, the Insular terrains, along with the Farewell and Arctic Alaska terrains, include crustal fragments that originated from separate sites within the Arctic realm in Paleozoic time.

Future allochthonous terrains may be found in the oceans, in the various plateaus, which are present on the ocean floor (Ben-Avraham et al., 1981). The plateaus are modern accreted terrains in migration, moving with the oceanic plates in which they are embedded and fated eventually to be accreted to continents adjacent to subduction zones. The present distribution of oceanic plateaus in the Pacific Ocean and the relative motion between the Pacific Plates and the Eurasian Plate suggest that the next large episode of continental growth will take place in the northwest Pacific margin, from China to Siberia. Similar situations may have occurred in the past in other parts of the world.

Thus, it may be possible to identify ancient accreted terrains in the geologic record, based on chemical and geologic evidence. The oldest preserved oceanic plateau sequences have been dated to 3.5 Ga in the Kaapvaal shield, South Africa (De Wit et al., 1987; Smith and Erlank, 1982), and the Pilbara Craton in Australia (Green et al., 2000). Greenstone belts of the Canadian Superior province, ranging in age from 3.0 to 2.7 Ga, also contain lava groups that have been interpreted to be remnants of accreted oceanic plateaus (see recent review by Kerr, 2013). The evidence for an oceanic plateau origin is based on the occurrence of pillow basalts and komatiites without terrestrial sedimentary intercalations or sheeted dyke swarms, possessing the characteristics of Cretaceous oceanic plateaus. Later in the text, we elaborate on the examples of crustal growth in the ANS and the Baltic Shield where evidence was found for the formation of oceanic plateau and accretion of juvenile terrains to the existing older continents.

9.07.4.3 The Transformation of Oceanic Plateau to Continental Crust

The accretion of oceanic plateaus with a clear oceanic affinity to continental margins is of fundamental importance in the mechanism of crustal growth. The crustal structure of these plateaus, which was originally composed of basalts, is modified with time and a typical structure of upper continental crust is developed. The most dramatic example of this process can be seen in British Columbia, northwestern Canada. Here, many accreted (or 'exotic') terrains were added to the continent during lithospheric plate convergence from the Archean to the present. The tectonic evolution of northwestern Canada involved a series of accretionary events, alternating with periods of continental extension. Thus, it is tempting to suggest that the continent has 'grown' westward by the surface area of the accreted terrains (Cook and Erdmer, 2005). The large-scale geophysical and geologic transect - "The Slave-Northern Cordillera Lithospheric Evolution" was carried out in northwestern Canada and showed quite clearly that several of the large accreted terrains, which have traveled long distances embedded within oceanic plates before colliding with the continent, have now a continental crustal structure (Cook and Erdmer, 2005). The Wrangelia and Stikinia terrains, which were accreted during the Phanerozoic, and Hottah terrain, which was accreted during the Proterozoic, have continental crustal structure with an upper crustal layer having granitic V_p velocities (Clowes et al., 2005).

Hollister and Andronicos (2006) proposed that crustal growth in the Coast Mountains, along the leading edge of the Canadian Cordillera, was the result of processes associated with horizontal flow of material during transpression and subsequent transtension and the vertical accretion of mantle-derived melts. Their model has two distinct tectonic phases. The first occurred during a period of transpression when the continental crust was thickened to about 55 km, and mafic lower crust of the Wrangellia oceanic plateau was pushed under the thickened crust where basalt from the mantle heated it to temperatures hot enough for melting. The basalt and the melts mixed and mingled and rose into the arc along transpressional shear zones, forming calc-alkaline plutons. The second phase occurred as the arc collapsed when a change of relative plate motions resulted in transtension that accompanied intrusion of voluminous calc-alkaline plutons and the exhumation of the core of the Coast Mountains batholith (**Figure 8(b)**).

The calculation made by Schubert and Sandwell (1989) suggests that accretion of all oceanic plateaus to the continents on a timescale of 100 Ma would result in a high rate of continental growth of 3.7 $km^3\ year^{-1}$. This rate is much higher than continental growth rate based only on accretion of island arcs, which is 1.1 $km^3 year^{-1}$ (Reymer and Schubert, 1986), thus providing an explanation to the previously mentioned crustal growth rate dilemma.

9.07.4.4 Mantle Overturn and Crustal Formation Episodes

In an attempt to link between various characteristics of mantle geochemistry and the dynamics and the apparent episodic mode of enhanced crustal growth (the Reymer and Schubert dilemma), Stein and Hofmann (1994) proposed that the geologic history of the mantle–crust system has alternated between two modes of mantle convection and dynamic evolution, one approximating a two-layer convective style when 'normal mode' of plate tectonics prevails (*Wilsonian tectonics*) and the other mode being characterized by significant exchange between the lower mantle and the upper mantle, when large plume heads form oceanic plateaus. The exchange periods apparently occurred over short time intervals – a few tens of million of years (as estimated from the chronology of magmatism that is involved in the production of the plume-related oceanic plateaus), while the Wilsonian periods extend over several hundreds of million of years (e.g., ~800 Ma in the case of the ANS; see later text). The idea is that during mantle overturn episodes (termed by Stein and Hofmann as *MOMO episodes*), a substantial amount of lower mantle material arrives to the Earth surface (mainly via plume activity), replenishes the upper mantle in trace elements, and forms new basaltic crust that eventually contributes juvenile material to the continental crust (via both subduction and accretion processes; **Figure 9**). Shortly after mantle upwelling and production of thick oceanic crust, new subduction zones are developed along the oceanic plateau margins as happened in the Mesozoic Pacific province. This mechanism views the formation of new subduction zones as a consequence of the production of thick oceanic crust and its cooling. Stein and Hofmann (1994) further proposed that

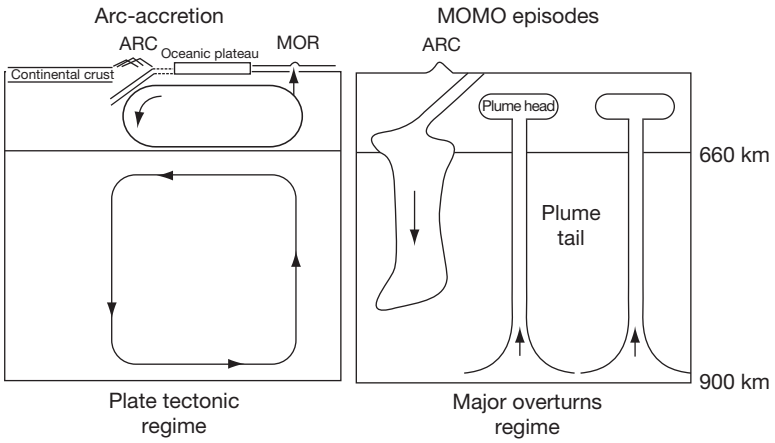


Figure 9 Mantle overturn and crust addition model (modified after the original MOMO model by Stein and Hofmann, 1994). The original MOMO model assumes that the mantle–crust system evolves through two modes of convection and dynamic evolution: one (MOMO overturn episodes) is characterized by significant exchange by the lower and upper mantle and crust and the other by two-layer convective style where typical plate tectonic Wilsonian cycle operates. The question of layered versus whole mantle convection has been continuously discussed in the literature and it is beyond the scope of this review. Here, we present a modified representation of the model where we stress the accretion of the oceanic plateau to the existing continent and development of subduction zones on the margins of the plateau (similar to the scenario shown by **Figure 8**). Most of the existing subduction zones in the world are associated with plateaus; a few subduction zones are not clearly associated with plateaus.

the mechanism responsible for the mantle overturns and rise of the plumes from the deep mantle is 'slab avalanches.' This mechanism describes the penetration of cold accumulated subducted material through the 660 km 'boundary' of the marginally stable two-layer convection configuration. It could also be that collapse of thick lithospheric roots that accumulated during the long processes of plate tectonics and subduction-related growth of continental crust (e.g., delamination or foundering models) could also trigger the collapse of the two-layer convection and rise of plumes from the deeper mantle.

The final product of the plume–thick oceanic crust-subduction system is a mass of continental crust that can be generated over a relatively short period of geologic time. The sequence of events involving the formation of anomalously thick, oceanic lithosphere, which is modified by subduction at its margin, explains the formation of both major continental segments during discrete episodes of time and their calc-alkaline affinities (a beautiful example for this evolution is described in the succeeding text for the growth of the late Proterozoic ANS).

In an alternative view of the Reymer and Schubert dilemma, Patchett and Chase (2002) and Patchett and Samson (2003) suggested that the evidence for enhanced crust production during discrete episodes can reflect the tectonics of accretion processes alone, arguing for the role of transform faulting, which serves to pile various terrains of juvenile crust into one restricted region (e.g., the Canadian–Alaskan Cordillera, where slices of juvenile accreted crustal material were piled up after northward along-margin transport associated with transform faulting; see **Figure 8**). The transport and accretion of juvenile mantle terrains are a central part of the plume-accretion model of crustal growth (**Figure 9**). Yet the additional important component in this model that is not required by the arc-accretion models alone is the evidence for the important role of oceanic plateau and enriched MORB crust as the nucleus of the juvenile continental crust and probably the locus of new subduction zones. Moreover, the crustal terrains that were formed during the episodes of 'major orogenic' production are not limited to geographically restricted region but rather are distributed globally. The best example is that of the late Proterozoic 'Pan-African' orogeny, when juvenile crustal terrains were produced in Arabia, Nubia (the ANS discussed later in the text) and North Africa, New Zealand, and Patagonia. All these geographic–crustal segments comprise the early Phanerozoic Gondwana. Considering a reconstruction of this old continent (**Figure 10**), it appears that continental segments within the Gondwana that can be identified as juvenile Pan-African crustal or lithospheric mantle terrains (see Stein, 2003, and caption of **Figure 10**) form a 'belt' along the possible margins of a mid or late Proterozoic continent. These segments are characterized by juvenile 'Pan-African' Nd, Sr, and Hf isotopic compositions that in turn are consistent with the plume ('MOMO')-type mantle (Stein and Hofmann, 1994, and see **Section 9.07.4.6**). In the following section, we discuss the isotope and chemical characteristics of the 'MOMO' mantle following the exciting recent ^{142}Nd data.

9.07.4.5 ^{142}Nd and Composition of the 'Juvenile Mantle'

The basaltic magmas that are the main samplers of the Earth mantle show no evidence for the existence of a 'primitive mantle' source. In this relation, it is interesting to return to the 'mantle array' that was first introduced by DePaolo and Wasserburg (D&W) (1976). D&W materialized the enormous potential in analyzing the ^{143}Nd/^{144}Nd and ^{147}Sm/^{144}Nd of various meteorites and determined these values for the chondritic meteorites, which were considered as samples of the 'primitive Earth' (the Chondritic Uniform Reservoir (CHUR) model) and then analyzed ^{143}Nd/^{144}Nd isotopic composition of oceanic and continental basalts and compared them to CHUR. Their achievement in the understanding of the

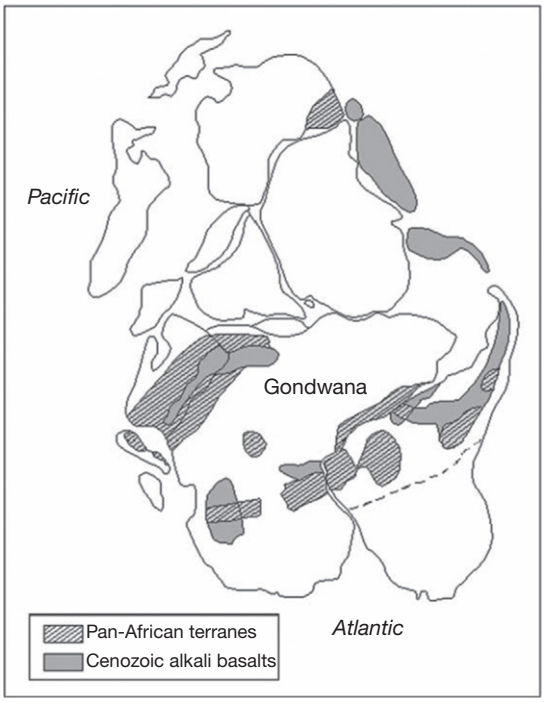


Figure 10 Distribution of regions that expose terrains of late Proterozoic Pan-African juvenile magmatic sequences and Cenozoic alkali basalts over a reconstructed map of Gondwana. The juvenile magmatic sequences are characterized by Nd and Sr isotopic compositions (e.g., $\epsilon Nd = +5 \pm 1$ and $^{87}Sr/^{86}Sr = 0.7028 \pm 2$; Stein, 2003; Stein and Hofmann, 1994) or Phanerozoic volcanic fields of alkali basalts (the Gondwana basalts over Africa, Arabia, Antarctica, and South America; the Pb isotopic compositions of these basalts are marked in **Figure 11** that also shows the plume–lithosphere connection) that were derived from the juvenile Pan-African lithospheric mantle. The areas that are best documented are those of the ANS on both sides of the current Red Sea (closed in the reconstruction). The figure suggests that during the late Proterozoic Pan-African orogeny, juvenile terrains were accreted to an existing continental nucleus in a similar manner to the growth that was described for the North American continent. The distribution of the juvenile Pan-Africa terrains may mark the locus of the subduction zones that were developed at that time along the margins of the Pan-African oceanic plateaus.

evolution of the Earth mantle and crust emerged from the properties of the Sm and Nd that are both refractory and lithophile elements (as all REEs) and were probably not fractionated during processes such as accretion and core formation and appear in nearly constant proportions in different groups of chondrites (e.g., Jacobsen and Wasserburg, 1980; Patchett et al., 2004) and thus were not easily fractionated by 'nebular' processes. Thus, the long ^{147}Sm–^{143}Nd (T1/2 = 106 Ga) system comprise a powerful tool for tracing the long-term chemical evolution of the silicate Earth. The original 'mantle array' of D&W comprised several basalts representing the major tectonic–magmatic environments: mid-ocean ridge basalt (MORB), Hawaiian basalts that represent the hot spot or plume-related OIBs, and the Columbia River Basalt (BCR, which has been extensively used as international standard for Sr and Nd isotope analyses) that represents continental basaltic magmatism. The Nd isotopic composition of the MORB yielded $\epsilon Nd = 10$, while the BCR sample yielded $\epsilon Nd = 0$, which according to the CHUR model is consistent with bulk silicate Earth. Based on the 'mantle array,' D&W proposed the 'simple model' for the structure of the Earth mantle where the primitive lower mantle (sampled by the BCR plume) is overlain by the depleted (in incompatible trace elements) mantle (sampled by MORB), and the complementary trace element-enriched reservoir would be the continental crust. It is important to note that the change (relative to CHUR) in Sm/Nd of the mantle could reflect the extraction of continental crust over the Earth history or instead was caused by early global differentiation of the Earth mantle such as may have occurred during an early formed magma ocean (see succeeding text).

The simple D&W model was later challenged by the overwhelming sets of Nd–Sr–Hf–Pb isotope data that indicated on a more complex mantle composition with at least four end-members that define the extended mantle array (see Hofmann, 1988). It should be stressed that the isotopic variability exhibited by the basalts indicates the long-term preservation (billions of years) of chemical heterogeneities in the convecting mantle (see reviews of this topic in Albarède and Van der Hilst, 2002; Hofmann, 1988; van Keken et al., 2002).

Important clues on the early history of the mantle isotope heterogeneities come from several extinct isotopes. A critical isotope pair in this respect is the relatively short-lived ^{146}Sm (T1/2 = 103 My) that decayed to ^{142}Nd in the early history of the Earth. The importance of the analyses of ^{142}Nd stems from its bearing on the $^{143}Nd/^{144}Nd$ ratio that reflects the fractionation of the $^{147}Sm/^{144}Nd$ in the Earth reservoirs (relative to the CHUR chondritic ratio). Over the past two decades, several workers made outstanding efforts to determine accurately the very tiny deviation of ^{142}Nd from CHUR.

Boyet and Carlson published a series of papers (e.g., Boyet and Carlson, 2005, 2006) showing a small, but significant, difference of ~20 ppm in the $^{142}Nd/^{144}Nd$ ratio between various types of meteorites and terrestrial samples. The deviation was identified in very old continental crust samples such as those from the Isua supracrustal belt (showing $^{142}Nd/^{144}Nd$ excesses of up to 17 ppm, e.g., Bizzarro et al., 2002). These results indicate that all terrestrial samples come from a mantle reservoir characterized by a superchondritic Sm/Nd ratio (e.g., **Figure 11**). The important implication of this result is that compared to chondritic meteorites (or the CHUR model), the elevated $^{142}Nd/^{144}Nd$ ratios of all terrestrial samples measured so far require that all these magmas were derived from a higher CHUR Sm/Nd ratio reservoir. This reservoir could reflect processes either within the solar nebula (namely, the excess in ^{142}Nd is not the result of ^{146}Sm decay) or in the mantle. The latter explanation implies that the silicate Earth has Sm/Nd ratio higher than chondrite due to an early differentiation (while ^{146}Sm was still present), during the first few hundred million years of solar system history to higher and lower than CHUR Sm/Nd reservoirs.

The main samplers of juvenile mantle material are the oceanic plateaus that according to the MOMO model represent plume head material that rose from the deep mantle and are manifested by thick sequences of tholeiites that typically comprise the early phase in the evolutionary cycle of juvenile provinces of continental crust (see succeeding text). **Figure 11**

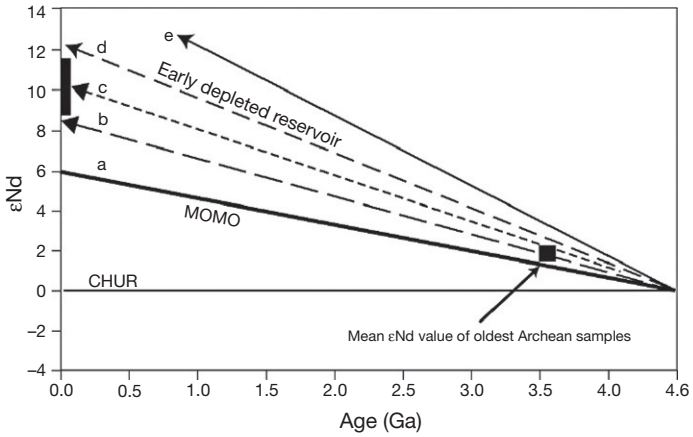


Figure 11 ϵNd evolution models of early depleted mantle. Line (a) comprises the 'MOMO–Nd evolution line' defined by the initial ϵNdt values of the major juvenile orogens (see Stein and Hofmann, 1994; **Figure 1**). The slope of the 'MOMO evolution line' corresponds to $^{147}Sm/^{144}Nd \sim 0.2$. Lines (b–e) represent $^{\epsilon}$Nd evolution lines that were calculated by Boyet and Carlson (2005), reflecting the excess of 20 ppm for the $^{142}Nd/^{144}Nd$ ratio in the Earth mantle relative to chondrites. Each evolution line is defined by a specific slope dependent on the magnitude of the Sm/Nd fractionation and the age of the differentiation event: (b) $\Delta t = 5$ Ma and the $^{147}Sm/^{144}Nd$ slope = 0.209; (c) $\Delta t = 30$ Ma and the $^{147}Sm/^{144}Nd$ slope = 0.216; (d) $\Delta t = 30$ Ma and the $^{147}Sm/^{144}Nd$ slope = 0.216; and (e) $\Delta t = 100$ Ma and the $^{147}Sm/^{144}Nd$ slope = 0.222. Evolution lines (c) and (d) lead to the range of ϵNd values in MORBs. The MOMO evolution line could be slightly rotated towards CHUR by mixing with an enriched early basaltic reservoir (following the Galer and Goldstein (1991) model).

shows that the MOMO evolution line of the juvenile orogens is consistent with Sm/Nd ratio that is higher than CHUR but lower than MORB-type mantle. The Sm/Nd ratio of the juvenile orogen evolution line is consistent with the estimate of Boyet and Carlson for the 'depleted mantle reservoir' after the early differentiation event that caused the ^{142}Nd excess (e.g., MOMO line (a) and the Boyet and Carlson calculated line (b) in **Figure 11**). The important implication of this result is that the magmas forming the juvenile orogens were derived from a depleted mantle that preserved its chemical identity since the early history of the Earth. In the context of the MOMO model, the plumes and the major juvenile orogens sample the 'early depleted mantle' that is currently characterized by $\epsilon Nd = +6$ and $^{147}Sm/^{144}Nd \sim 0.2$. This mantle is represented by intraplate continental basalts that are erupted in all continents and was acronymed by Wörner et al. (1986) as the PREMA mantle reservoir or as the FOZO mantle reservoir (Hart et al., 1992). We prefer the acronym *MOMO* that provided also the geodynamics and the Earth history perspective of the term. As the MORB represents the depleted uppermost mantle and its tectonic environment, MOMO represents the connection to the mantle plumes and major orogens that were primarily evolved from the 'early depleted mantle.'

9.07.4.6 Hf in Zircons and Crust Recycling

The extensive works that were performed on zircons from granitoid terrains and detrital zircons (e.g., in Phanerozoic clastic (sandstones) sequences) provide rich data that are complementary to the information derived from Nd and Sr isotopes from the bulk magmatic rocks. Oxygen isotope ratios derived from the zircons are used to identify those that were crystallized in juvenile mantle magmas (e.g., 'mantle-type' $\delta^{18}O$ values of ~+6‰) or show significant recycled crust that contained surface sediments or was affected by seawater (e.g., $\delta^{18}O$ values higher than +7‰). Many of the initial ϵHf(t) detrital zircons and some of the granitoid zircon data lie below the 'primitive mantle' evolution line in **Figure 12** (detrital zircons that spread to low ϵHf(t) values (e.g., −10 to −30 during the past 2 Ga) are not plotted and are presented in several recent works, e.g., Iizuka et al., 2013), indicating contribution of recycled crustal material to the magmas. Yet a large number of the granitoid zircons and the detrital zircons converge to values that lie between the 'primitive mantle' and 'depleted MORB mantle' (**Figure 12**). They lie on an Hf evolution line defined by the 'juvenile orogens' similar in its concept to the 'MOMO–Nd evolution line' that was described for the initial Nd isotopes of the 'juvenile orogens.' The extensive set of oxygen and zircon isotope data from late Proterozoic ANS granitoids that are exposed in the Sinai Peninsula allows us to plot the field of the juvenile ANS on the 'MOMO–Hf evolution line.' The ANS–MOMO–Hf field is defined according to zircons with mantle-like $\delta^{18}O$ values. Similarly, Stein and Hofmann (1994) defined the MOMO–Nd evolution line of the juvenile orogens according to the $^{87}Sr/^{86}Sr$ and $\delta^{18}O$ values that were available for the magmas. We note that again similar to the 'MOMO–Nd evolution line,' the 'MOMO–Hf' evolution line is very close to the line described for subduction zones (see the 'arc mantle' in Dhuime et al., 2011). This reflects the composition of the arc volcanics that were derived from the MOMO mantle during the long interval of subduction-related magmatism that processed the MOMO – mantle to calc-alkaline magmas that comprise substantial parts of the upper continental crust.

9.07.5 The Production of Calc-Alkaline and Alkaline Granites and Related Rocks

9.07.5.1 General

The average composition of the upper continental crust is granodioritic (Taylor and McLennan, 1985), reflecting the

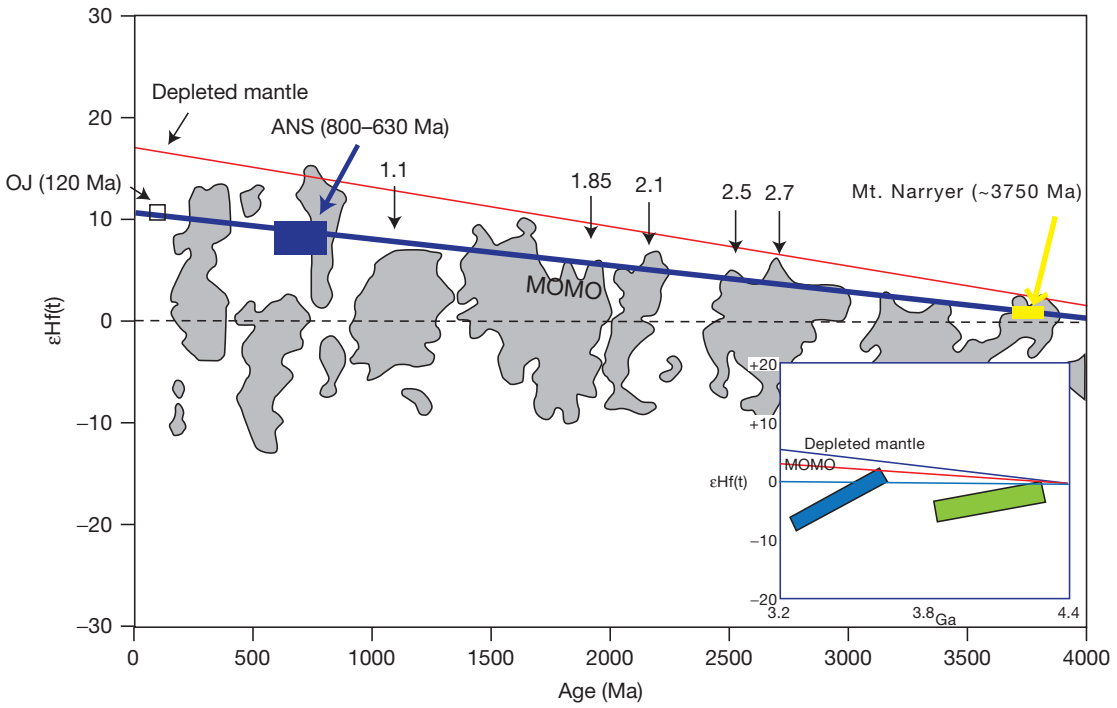


Figure 12 ϵHf isotope evolution diagram of primitive mantle (ϵHf = 0), depleted mantle, and major continental crustal segments (the 'major orogens'). The diagram shows Hf isotope data of zircons from granitoid terrains against their U–Pb crystallization ages (reproduced from Guitreau M, Blichert-Toft J, Martin H, Mojzsis SJ, and Albarede F (2012) Hafnium isotope evidence from Archean granitic rocks for deep-mantle origin of continental crust. *Earth and Planetary Science Letters*. 337/338:211–223; Arndt NT (2013) Formation and Evolution of the continental crust. In: Elliott T (ed.) *Geochemical perspectives*, vol. 2, No 3, and references there). The ages of the 'major orogens' are marked. The MOMO evolution line is defined by eNd(t) values of Ontong Java basalts (OJ) (data from Tejada et al., 2004), zircons from the juvenile ANS granitoids that are filtered by their mantle-type $\delta^{18}O$ values (data from Morag et al., 2011) and the Hadean Mount Narryer zircons (see text). Most of the Hf zircon data form fields that lie below the MOMO evolution line marking crustal recycling. The inset shows the ϵHf(t) values in early earth zircons (reproduced from Bell EH, Harrison TM, McCulloch MT, and Young ED (2011) Early Archean crustal evolution of the Jack Hills zircons. *Geochimica et Cosmochimica Acta* 75: 4816–4829). See text for further discussion of these data.

important role that calc-alkaline granitic magmas take in the construction of the continental crust besides the subduction-related basaltic–andesitic magmas and the tholeiitic oceanic plateau-type basalts. Yet the petrogenetic conditions and geodynamic relation of the calc-alkaline granites still comprise open issues, and surprisingly, a limited number of geochemical–petrological works were devoted to the understanding of the mechanism and geodynamic environment of production and emplacement of the granitic magmas (e.g., Borg and Clynne, 1998; Chappell et al., 1987; Green and Ringwood, 1968; Smithies et al., 2011; Wyllie, 1984; Wyllie et al., 1976). Unresolved issues include the composition and location of the sources of the calc-alkaline magmas, the sources of heat that are required for the melting, the mechanism of fractionation that forms the calc-alkaline series (e.g., from gabbro, diorite, tonalite, to granodiorite and granite), the mechanism of emplacement of vast amounts of granitic plutons, and their place in the chronology of events during the formation of juvenile segments of continental crust. This section discusses several topics in the broad field of petrogenesis of calc-alkaline (I-type) and alkaline (A-type) granites and related rocks. Here, we describe two recent works on the petrogenesis of calc-alkaline granites from the Elat massif (Weismann et al., 2013) and A-type granites (and other magmas) from the Amram massif (Mushkin et al., 2003). Both massifs are located at the northern uplifted edge of the ANS. The big advantage and beauty of exploring the calc-alkaline and alkaline massifs at Sinai Peninsula and Elat area lie in their unique 'desert clean and fresh' exposures and the variety of the rock types from mafic to felsic. The massifs were uplifted 'recently' during the Miocene to Pleistocene tectonic events that are associated with the breakup of the African plate and opening of the Red Sea and formation of the Suez Rift and Dead Sea Transform. The Elat massif comprises felsic calc-alkaline rocks, mafic xenoliths, and intermediate host rocks that were all emplaced during the calc-alkaline phase of the ANS history ~630–600 Ma (Weismann et al., 2013, and references on the chronology there). This association in space and time is supported by their systematic and continuous variation in major element concentrations and their similar initial Sr isotope ratios, suggesting a genetic link between all magma types. In the section later in the text, we discuss several models that were attributed to explain the petrogenesis of calc-alkaline magmas and apply the results from the Elat massif suite to infer the model of mafic crustal anatexis.

9.07.5.2 Models of Calc-Alkaline Granite Production

Three general processes can be considered for the production of this calc-alkaline association: (1) fractional crystallization of

mantle-derived magma similar in composition to the mafic xenoliths towards the formation of the intermediate and felsic host rocks, with possible involvement of crustal assimilation; (2) mixing between mafic and felsic magmas; and (3) melting of crustal rocks by emplacement of mantle-derived basaltic magma in the presence of various amounts of water.

Fractional crystallization of a mantle-derived magma similar in composition to the mafic xenoliths requires high degrees of fractionation, that is, only small volumes of felsic rocks will form from a given volume of mafic magma in such a process. It is unlikely that such a mechanism produced such a large volume of felsic plutons that comprise most of the exposed surface of the northern segments of the ANS. In several cases, fractional crystallization accompanied by assimilation of crustal rocks (AFC, assimilation–fractional crystallization) is the suggested mechanism for the formation of I-type granites (DePaolo, 1981). This model implies that mafic magma, sourced from the mantle, is modified during its ascend and emplacement at crustal levels due to assimilation of crustal material during fractional crystallization, leading to the formation of felsic rocks. The heat required for assimilation of the crustal rocks is provided by the latent heat of crystallization of the magma.

9.07.5.3 The Petrogenesis of Calc-Alkaline Granites and Related Rocks from the Elat Massif

The calc-alkaline granites in the Elat area were indeed emplaced into metasedimentary rocks (e.g., the Elat schist), which could be regarded as the possible crustal component assimilating the primitive basaltic magma. Bielski (1982) reported initial $^{87}Sr/^{86}Sr$ ratio (ISr) of 0.7030 ± 6 for the Elat schist, suggesting that these rocks and the calc-alkaline granitic suite (ISr value of 0.7034; Weismann et al., 2013) were derived from young crustal rocks. Be'eri-Shlevin et al. (2010) performed a multi-isotope study of the calc-alkaline phase of the ANS in the Elat area and observed high $\delta^{18}O$ values for some of the Elat calc-alkaline granites, associated with initial ϵNd (T) lower than the 'juvenile values' (inferred as ~+4; Stein and Goldstein, 1996) that were uncorrelated with variations in the major element chemical composition. Such a correlation would have been expected if AFC process had been a main mechanism associated with the formation of the calc-alkaline granites.

Basaltic mantle-derived melts ascending to crustal depths can induce partial melting of crustal rocks ('crustal anatexis') of felsic compositions, leaving behind a hydrous (amphibole-bearing) mafic residual rock (e.g., Bergantz, 1989; Droop et al., 2003; Huppert and Sparks, 1988; Sparks et al., 1977; Wyllie et al., 1976; Wyllie, 1977, 1984). The partial melting of the crust is enabled due to the heat and water supplied by the intruding basaltic melt. Various degrees of mixing of the mantle-derived mafic magma and the crustal felsic magma can lead to a wide range of magma compositions. Mafic xenoliths or xenocrysts surrounded by felsic magmas could be the products of such 'mingling.'

A common characteristic of calc-alkaline granitic plutons formed from magma mixing is the production of chemical zoning within the plutons (Zorpi et al., 1989). Since all mafic and felsic calc-alkaline rocks in the Elat massif lie on the same isochron converging to low ISr value of 0.7034, different sources (which evolved separately) for the mafic and felsic magmas are less likely. If mixing or mingling mechanism led to the formation of all calc-alkaline rocks in the area and mafic magmas intruded a felsic magmatic cell, both magmas must have been derived from the same source. Radiogenic isotope ratios (Nd–Sr–Pb) from the ANS core region were previously interpreted to show limited involvement of pre-Neoproterozoic crust in its formation (Be'eri-Shlevin et al., 2009, 2010; Stein, 2003; Stein and Goldstein, 1996).

Several studies advocated the 'restite unmixing model' to explain the production of suites of rocks of felsic compositions (e.g., Clemens, 1989; White and Chappell, 1977; Wyborn and Chappell, 1986). In this model, the rocks represent various degrees of separation between the residual source material and the melting phase. A possible petrogenesis scenario of Elat massif includes melting of the continental crust composed of rocks similar to Roded quartz diorite. The calc-alkaline rocks represent various degrees of melt separation (leading to the formation of the intermediate to felsic magmas) from the residual source (mafic xenoliths). The lack of linear correlation among the trace elements of the Elat suite and the common poikilitic texture argue against the restite unmixing model and rather indicate the formation of the calc-alkaline suite via crystallization from melt.

9.07.5.4 Crustal Anatexis and the Role of Water

Experimental studies have shown that water has a profound effect on melt compositions and residuum mineralogy during partial melting of mafic rocks (e.g., Beard and Lofgren, 1989; Holloway and Burnham, 1972; Rushmer, 1991, Wolf and Wyllie, 1994). In this model, crystallization of mantle-derived basaltic magma near the base of the colder crust provides the heat and, possibly, the hydrous fluids required for crustal anatexis. The introduction of water and heat leads to melting of the crustal rocks, but differences in fH_2O can account for differences in the resultant magma compositions and residual rock. In cases where the only source of water is the breakdown of hydrous minerals in the crustal rocks (dehydration melting), a tonalitic melt composition is formed, leaving behind a dry residuum composed of plagioclase, pyroxene, and Fe–Ti oxides (Beard and Lofgren, 1989; Wolf and Wyllie, 1994). When water is available as an external phase, vapor-saturated melting of the same crustal rock produces high silica and alumina felsic melts coexisting with an amphibole- and plagioclase-bearing restite assemblage (Beard and Lofgren, 1989; Helz, 1976).

The evolution of continental crust in the ANS spanned more than 250 Ma of subduction-related calc-alkaline volcanism before the time of production of the calc-alkaline granites and diorites. The subduction-related volcanics comprised the protolites of the thick sequences of metavolcanic and metapelitic rocks. The Elat amphibolite exposed in the Elat massif could represent the potential source for the calc-alkaline granites. The hydrous–basaltic protolite of the amphibolite was produced during the calc-alkaline–arc volcanic phase and then underwent metamorphism. The high and uniform values of amphibole/melt partition coefficient for many incompatible trace elements imply that melting of amphibolitic source could produce the calc-alkaline granites with similar distribution of incompatible trace elements between the amphibolite and the

magma. The hydrous–basaltic protolites of the amphibolites could penetrate into the lower crust and be metamorphosed there. *Intruding hydrous basalts could serve as a source for heat and water for partial melting of former amphibolitic rocks and the production of felsic magma*. The high-water content in the amphibolitic rocks resulted in vapor-saturated melting, producing felsic melts, as represented by Elat calc-alkaline granite, and an amphibole-rich residuum, as represented by the mafic xenoliths.

9.07.5.5 The Genesis of Alkaline A-Type Granites

Anorogenic alkaline granites (termed A-type granites) and related A-type alkaline magmas commonly mark the closing stages of a major orogeny and are related to postcollision and intraplate tectonic setting (e.g., Black and Liegeois, 1993; Eby, 1990; Turner et al., 1992). Several mechanisms were proposed for the formation of the A-type granites: (1) partial melting of crustal rocks (e.g., Collins et al., 1982; Creaser et al., 1991; Whalen et al., 1987) and (2) fractionation products from the differentiation of mantle-derived mafic magmas (e.g., Kessel et al., 1998; Mushkin et al., 2003; Stern and Gottfried, 1986; Turner et al., 1992; Volkert et al., 2000). Here, we briefly describe the work of Mushkin et al. (2003), who studied the beautifully exposed Amram massif, north to the Elat massif. The Amram massif comprises a suite of mafic to felsic A-type magmas: rhyolites, alkali quartz syenites, quartz syenites, monzonites, and comagmatic mafic to felsic alkaline dikes (45.6–78.8% SiO_2). The rocks were formed ~550–530 Ma, upon the closing stage of the Pan-African orogeny at the ANS and the transition from an orogenic to an intraplate tectonic setting in the northeastern ANS, and reveal a correlation between decreasing stratigraphic age and increasing silica content. The Amram mafic rocks ($SiO_2 = 45.6–49.5\%$) have high MgO and Fe_2O_3 concentrations (4.10–8.95% and 10.0–12.5%, respectively) and relatively flat REE patterns $(La/Yb)_N = 6.4 \pm 0.9$, suggesting derivation from relatively primitive, mantle-derived magmas. It is proposed that the mafic magmas were derived from the juvenile lithospheric mantle, underlying the ANS. The Amram mafic and felsic rocks define a continuous chemical evolutionary trend and display constant La/Nb and Y/Nb ratios (1.4 ± 0.3 and 1.6 ± 0.3, respectively). Hence, the evolution of the Amram suite is attributed to fractionation of parental mafic magmas with only minor incorporation of crustal material. Thermodynamic modeling (applying MELTS) relates the Amram high silica (>70%) rhyolites to extensive (>90%) fractionation under anhydrous conditions, with plagioclase, alkali feldspar, clinopyroxene, olivine, and minor Ti magnetite and apatite as the fractionating phases. This fractionation would require a much larger volume of unexposed cumulate rocks (composed of the fractionating mineral phases) and would imply that the parental mafic magmas were derived from a large source region in the ANS lithospheric mantle. The post-orogenic alkali magmatism represents juvenile addition to the upper continental crust, of a similar magnitude to that of the Cenozoic rift-related alkaline magmatism of the Arabian Plate (e.g., Weinstein et al., 2006). Yet all these alkali basalts are possibly derived from lithospheric mantle sources that were primarily formed during the late Proterozoic orogeny in the ANS. Thus, the 'net juvenile' contribution to the Arabian continent is small (see succeeding text).

9.07.5.6 Petrogenesis of Granites: Concluding Comments

The described studies on the petrogenesis of the late Proterozoic calc-alkaline and alkaline granites from the northern ANS put them in the context of the evolutionary cycle of the ANS continental crust. It was the long interval of subduction-related magmatism and arc volcanics (250 Ma, comparable to the duration of activity of subduction processes along the Japanese arc) that produced the juvenile addition to the new continent, leading to the accumulation of thick lithosphere and crust that eventually went through anatexis of the crust or melting of the subcontinental lithosphere and not only produced a significant amount of felsic magmas but also left behind residual lithologies that could be dropped back to the mantle. The previously described models of crustal anatexis or lithospheric mantle melting that produced the ANS granitic magmas place the 'granite factory' in the 'intraplate' tectonic environment. This is simply because subduction processes in the ANS ceased before the late orogenic and post-orogenic production of the felsic granitic batholiths. Yet granitic magmas can be produced in the subduction environment itself. The 'subduction factory' provides the basic conditions for granite formation: parental–basaltic magmas and presence of waters and accommodates processes of fluids release from the subducted slab due to the breakdown of hydrous minerals (e.g., Grove et al., 2006), production of hydrous–basaltic magmas from the melting of hot peridotite + fluids, and fractionation of the hydrous–basaltic magmas to produce tonalitic and eventually granitic magmas. Regarding the history of the ANS, granitic metagneisses are known from the early phase of subduction-related magmatism in the ANS (e.g., Bentor, 1985). Nevertheless, most of the granitic batholiths covering the northern ANS and granitic batholiths in other major juvenile crustal orogens (see succeeding text) were produced after the termination of subduction activity (see succeeding text).

The thick sequence of 'meta-arc' volcanics including hydrous–basaltic magmas that accumulated during subduction period comprises the potential source for the calc-alkaline granites. We reemphasize the critical role of waters in the 'granite factory.' All models that propose the production of granites from thick basaltic sequences require the supply of water to the melting environment (no water, no granites). This prime requirement of the presence of water in the production environment of granites brings us back to the role of oceanic plateaus in the growth of continental crust. Several workers suggested that the thick piles of Archean basaltic sequences (e.g., original plume-related picritic magmas) comprising the basal parts of oceanic plateau sequences were melted to produce crustal tonalitic–granitic magmas (e.g., Bédard, 2006; Van Kranendonk, 2010). The problem is to bring water to the basal parts of the plateaus, which appear to be dry (Arndt, 2013, and references there). Arndt argued that only the uppermost part (1–5 km) of the layered plateau sequences comprises lavas that could interact with seawater. The basal parts of the plateau comprise crystallized phases from the primary plume-related picritic melts and compact gabbroic layers that are less susceptible to water penetration and alteration. In this context, we return to the interesting

earlier-mentioned observation from the Ontong Java Plateau where crustal seismic velocities were identified in the upper 5–15 km (Figure 7). This issue remains an open question that requires further attention.

9.07.6 Rapid Growth of Major Continental Segments (the 'Major Orogens')

9.07.6.1 General

In this section, we describe of the magmatic histories of several prominent crustal segments (termed as 'major orogens') that underline the history of the Earth continental crust. We first elaborate on the evolution of continental crust comprising the late Proterozoic ANS. Then, we describe the other major orogens. Overall, we show the general similarity in the evolutionary pattern of all these orogenic episodes. They all evolved through four major phases: the (I) eruption of thick sequences of tholeiitic magmas. (II) production of arc subduction-related calc-alkaline magmas, (III) production of calc-alkaline granitic batholiths, and (IV) post-orogenic production of alkaline (A-type) granitoids.

9.07.6.2 The ANS Orogeny (~0.9–0.6 Ga)

The excellent preserved exposures of the late Proterozoic crystalline basement of the ANS and the sequences of overlying Phanerozoic alkali basalts provide an opportunity to monitor the magmatic history of the ANS over the past ~900 Ma (Stein and Goldstein, 1996). The low initial $^{87}Sr/^{86}Sr$ ratios of the late Proterozoic magmas of Sinai, Jordan, northern Saudi Arabia, and the Eastern Desert of Egypt preclude significant involvement of an old pre-Pan-African felsic continental precursor in this part of the northern ANS, indicating that basement magmas were derived either from a preexisting mafic crust or directly from the mantle. The absence of older continental crust is further indicated by mantle-type oxygen isotope ratios (e.g., $\delta^{18}O = +5.5$, Bielski, 1982). Recent studies of 'magmatic detrital' zircons from the ANS granitoid plutons and the Phanerozoic sedimentary cover, respectively, confirmed the juvenile nature of significant areas of the ANS (e.g., Be'eri-Shlevin et al., 2009, 2011; Morag et al., 2011).

The late Proterozoic history of the ANS can be divided into four evolutionary phases (following Bentor, 1985):

Phase I: This phase is characterized by the production of thick sequences of tholeiitic during a very short interval of time (~900 ± 30 Ma). The (meta)tholeiites are characterized by chondritic REE patterns and 'mantle-type' Th/Nb ratios and were interpreted as oceanic basalts similar to the plateau basalts (Stein and Goldstein, 1996; Stein, 2003, and references within there). It is important to emphasize that along with the production of the plateau magmas, thinner sequences of metatholeiites were interpreted as 'normal' mid-oceanic tholeiites erupting along spreading ridges (the 'old metavolcanics' in Egypt, described by Stern (1981)). The two ANS metatholeiite suites resemble the magmatic–tectonic association of the Pacific Ontong Java Plateau–Nauru Basin basalts.

Phase II: Upon reaching a continental margin, the thick oceanic plateau lithosphere resisted subduction and plate convergence occurred on its margins, forming the ANS calc-alkaline magmas. The subduction activity lasted for >200 My (between ~870 and 650 Ma), similar in duration to the activity arc magmatism in the Japanese arc. The large amount of calc-alkaline magmatism over a short period of time may be partly a result of the thermal anomaly in the upper mantle caused by the plume head. The arcs and oceanic plateaus were accreted to the old Gondwanaland producing thick and mainly mafic crust and lithospheric mantle, which served later as sources for the calc-alkaline granitic batholiths, alkali granites, and Phanerozoic alkali basalts. The enriched plume–mantle of the early ANS was exhausted by the subduction processes. The Gabal Gerf ophiolites (~750 Ma; Zimmer et al., 1995), which are assumed to represent the arc environment, show isotope mixtures between the enriched ('plume–lithosphere mantle') and depleted ('MORB-type mantle') components (see evolution lines in Figure 13). A similar shift to more depleted

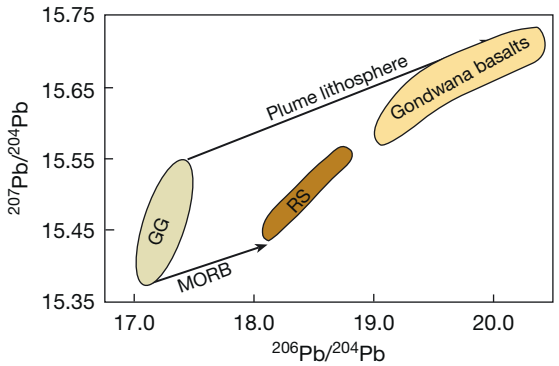


Figure 13 The distribution of late Proterozoic oceanic basalts (the Gabal Gerf (GG) ophiolites, Red Sea (RS) basalts, and Gondwana basalts in the Pb–Pb isotope diagram (reproduced from Stein M and Goldstein SL (1996)) From plume head to continental lithosphere in the Arabian-Nubian shield. *Nature* 382: 773–778; Stein M (2003) Tracing the plume material in the Arabian-Nubian Shield. *Precambrian Research* 123: 223–234). Stein and Goldstein proposed that the GG ophiolites lie in the diagram on a mixing trend between the asthenospheric MORB mantle and asthenospheric plume-type mantle (in similar tectonic–magmatic settings to the configuration of the Nauru Basin and EPR in the Pacific Ocean). The Gondwana basalts (location marked in **Figure 10**) lie on an evolution trend together with the late Proterozoic-enriched ophiolites and other magmas (e.g., galenas and K-feldspars not shown here; see Stein and Goldstein, 1996), suggesting that the lithospheric mantle beneath the Arabian continent was generated during the late Proterozoic crustal formation events. It can be regarded as a 'frozen' mantle wedge, which previously accommodated the production of calc-alkaline magmas and fractionation–transportation processes of trace elements on route to shallower levels of the crust (the subject is elaborated later in the text). The Red Sea basalts lie on a mixing trend between the lithospheric-derived basalts and the MORB-type asthenosphere. Stein and Hofmann (1992) noted that enriched-type Red Sea basalts erupt on the southern and northern edges of the Red Sea basin, while depleted MORBs erupt only in the central trough where 'real' spreading occurs. They regard this configuration as a strong evidence for the existence of the plume-type mantle above the asthenospheric MORB. The latter is rising into the rifting lithosphere and continental crust. This also provides a strong evidence for the MORB-type composition of the upper mantle. The Gondwana basalts show mixing relation with the Red Sea asthenosphere, suggesting that MORB is rising into the rifting lithosphere (e.g., in the Ross Sea rift, Ahaggar, and the Dead Sea rift). This rising has important implications for the heating of the lithosphere and the lower crust and for the production of basaltic melts from the lithosphere and provides evidence for asthenospheric underplating (Stein, 2006).

MORB values is observed also in Hf isotopes (eHf >+ 10) in zircons recovered from the Pan African magmas (Fig. 12)

Phase III: This phase constitutes the rapid production (~630–600 Ma; Eyal et al., 2010) of vast amounts of the calc-alkaline granitic and dioritic magmas, which form batholiths that cover large areas, mainly in the northern ANS. The petrogenesis of the calc-alkaline granitoids (**Section 9.07.5.2**) was related by Weismann et al. (2013) to anatexis of metamorphosed mafic crust (e.g., the Elat amphibolite).

The production of the calc-alkaline granitic magmas certainly required the supply of 'heat and water,' a problem that is not clearly understood. One possibility is a relation between asthenospheric rise associated with delamination of a thick lower crust and lithosphere that were produced in the earlier evolutionary stage of the island arc magmatism.

Phase IV: This phase is represented by the orogenic cycle terminated by the production of alkali granites and related rhyolites (~600–530 Ma; Bielski, 1982; Eyal et al., 2004; Mushkin et al., 2003; Morag et al., 2011). The parental magmas of the alkaline magmas were probably derived from the lithospheric mantle and underwent extensive differentiation in shallow crustal levels (Mushkin et al., 2003).

The transition from calc-alkaline magmatism to alkaline magmatism associated with the lithospheric mantle melting after 600 Ma reflects the ending of plate convergence. In the ANS, this transition is associated with a major uplift (Black and Liegeois, 1993). Similar rapid transitions in the stress field associated with continental collisions were described in other localities such as the Variscan province in Europe, the Basin and Range Province in the United States, and the Tibetan Plateau (Costa and Rey, 1995; Klemperer et al., 1986; Rey, 1993). In the Himalayas, major uplift and the appearance of alkaline volcanism have been related to the thinning and melting of the lithosphere after the ending of plate convergence.

The growth and fate of continental crust in the ANS are schematically illustrated in **Figure 14**. Overall, most of addition juvenile material to the ANS continental crust occurred during phases I and II during its early 200–250 Ma of plateau and arc formation and accretion. Magmas formed during phases III and IV comprise melting products of crust and lithospheric mantle that were formed in the earlier phases. Thus, the rate of juvenile creation of crust in the ANS (and possibly other 'major orogens') could be even faster than the Reymer and Schubert estimate.

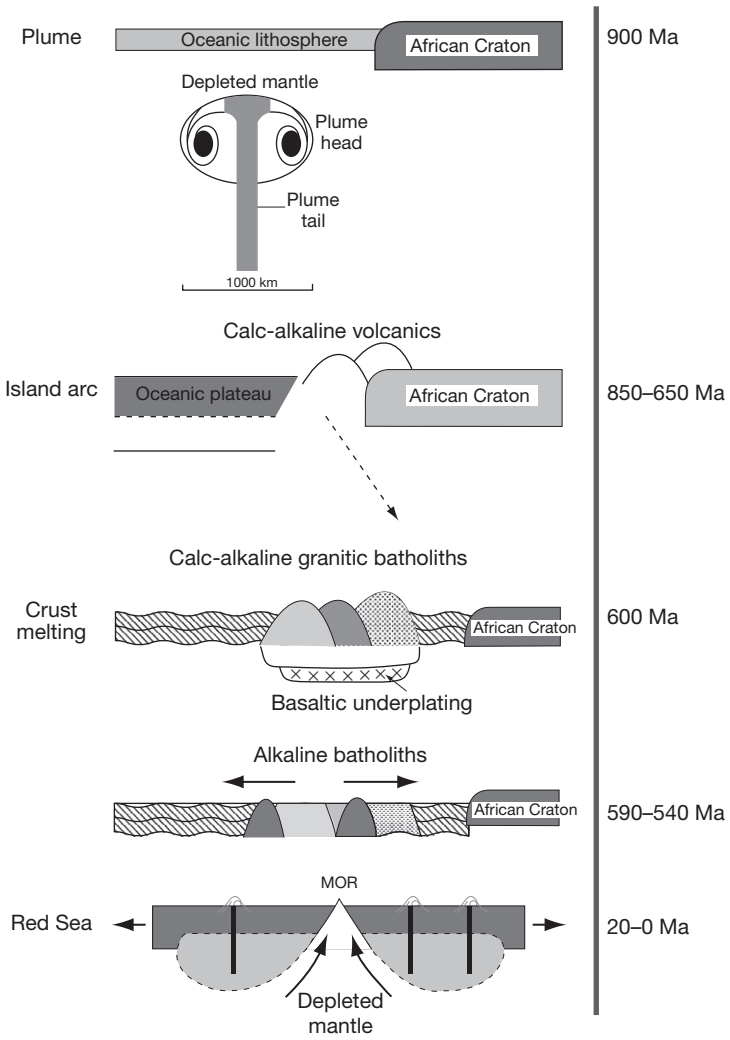


Figure 14 Schematic model illustrating the growth and fate of continental crust and lithosphere in the ANS. Similar (but clearly not identical) structure of magmatic histories can be ascribed to other major crustal segments (e.g., Canadian Shield and Birimian orogen the Mesozoic Pacific).

9.07.6.3 The Birimian Orogeny (~2.2–1.8 Ga)

Another well-studied case of episodic crustal formation is that of the Birimian orogeny that lasted between ~2.2 and ~1.8 Ga (Abouchami et al., 1990; Boher et al., 1992). The early stage in the growth of the Birimian juvenile crust was marked by eruption of thick sequence of basalts in oceanic basin that were interpreted as oceanic plateau (formed in an environment remote from any African Archean crust). Shortly after, calc-alkaline magmas were produced, probably in island arcs that were developed on the margins of the oceanic plateau, which then collided with the Man Archean craton.

While the magmatic–geodynamic history of the Birimian orogeny strikingly resembles that of the ANS and to some extent that of the Superior province, the growth rate calculated for the Birimian crust is ~1.6 $km^3 a^{-1}$, which is only ~60% higher than the Phanerozoic current rate (Reymer and Schubert, 1984) and slower than the mean growth rate of 2 $km^3 a^{-1}$ of the total continental crust from the mantle in 4 Ga (Turcotte and Schubert, 1982). Boher et al. argued that large part of the Birimian crust has been recycled into the mantle or incorporated into younger orogenic segments. They extended this line of argument to the Archean crust, suggesting that the apparent deficit in crustal budget is even more dramatic for the Archean crust. Thus, while the evidence from Birimian orogeny seems to support the model of episodic formation of juvenile continental crust in relation to large mantle upwellings, it also outlines the importance of recycling processes.

9.07.6.4 The Superior Orogeny (~2.7–2.5 Ga)

The Superior province, the largest Archean craton on Earth, comprises several continental and oceanic megaterrains that were formed and accreted rapidly (the ages of the various superprovinces lie in the range of 3.1–2.67 Ga and accretion–tectonics occurred between 2.74 and 2.65 Ga; Card, 1990; Percival et al., 1994).

Recent compilations of U–Pb zircon ages (e.g., Condie et al., 2009, 2011; **Figure 12**) indicate that the 2.7 Ga peak is the largest one and reflects a global event that was recorded on all continents: for example, Canada (e.g., Card (1990)), Zimbabwe (e.g., Bickle and Nisbet, 1993), Siberia (e.g., Puchtel et al., 1993), and Australia (e.g., Myers and Swagers, 1997), a picture that is consistent with the concept of the MOMO model on a global overturn event in the mantle.

The geologic history and magmatic–stratigraphy of the Superior province were divided by Arth and Hanson (1975) into four stages: (1) the production of tholeiitic basalts by shallow melting of mantle peridotites, followed by fractional crystallization that produced basaltic–andesites; (2) the formation of trondhjemite, tonalite, and dacites by melting of eclogites in mantle depth; (3) the intrusion of quartz monzonites; and (4) the intrusion of postkinematic syenites. The general pattern of this history is strikingly similar to the four-phase growth and magmatic evolution of the previously described late Proterozoic ANS or the Proterozoic Birimian orogeny.

In more recent studies, Polat and Kerrich (2001a,b) and Polat et al. (1998) described the formation of the 2.7 Ga Wawa and Abitibi greenstone subprovinces of the Superior province as an example of episodic continental growth, by lateral spread of subduction–accretion complexes. They recognized two principal volcanic associations: (1) tholeiitic basalt–komatiite and (2) tholeiitic to calc-alkaline bimodal basalt–rhyolite. Tholeiitic basalts of the former association are characterized by near-flat REE patterns and Th/U, Nb/Th, Nb/La, and La/Sm-pm ratios that span the primitive mantle values. Transitional to alkaline basalts and Al-depleted komatiites have fractionated REE patterns and OIB-like trace element signatures. Polat and his coauthors interpreted these assemblages as representing oceanic plateau derived from a heterogeneous mantle plume. The bimodal association has fractionated REE patterns and negative Nb, Ta, P, and Ti anomalies typical of arc magmas. Tonalite plutons derived from partial melting of subducted oceanic slabs intrude the subduction–accretion complex as the magmatic arc axis migrated towards the trench. The eruption of voluminous ocean plateau and island arc volcanic sequences and the following deposition of kilometer-thick siliciclastic trench turbidites and intrusion of the supracrustal units by syn- to postkinematic plutons occurred in the relatively short period of time of 2.75–2.65 Ga (a classic representation of the Reymer and Schubert superrapid crustal formation events). The earlier-mentioned events were accompanied by contemporaneous intense polyphase deformation, regional greenschist to amphibolite facies metamorphism, and terrain accretion along a north–northwest dipping subduction zone.

The description of the history of the Superior province is well fitted with the mantle plume upwelling–arc magmatism and terrain accretion models where the growth of the new continent and lithosphere involves accretion of oceanic plateaus, island arcs, continental fragments, closure of oceanic basins, rifting of magmatic arcs, and plume–arc interactions followed by intracrustal and intralithospheric extension-related melting. This comprises a complete MOMO–Wilsonian cycle that was repeated throughout the Earth history.

9.07.6.5 Early Precambrian Orogenies (~2.9–2.8 Ga)

The previous sections suggest that the crustal growth histories of the ANS and apparently the Superior and Birimian orogenies comprise complete MOMO cycles, illustrating a primary means of continental crustal formation and cratonization throughout the Earth history, and were particularly relevant in cases of rapid continental growth as demanded by the Reymer and Schubert dilemma. Evidence for the existence of magmatic provinces that resemble oceanic plateaus was found in older continental crust provinces.

The 2.9 Ga Sumozero–Kenozero greenstone belt in the SE Baltic Shield comprises several kilometer-thick lower unit that is made of submarine mafic–ultramafic units (with plume-type Nb/U ratios of 43 ± 6) that were interpreted as representing oceanic plateau magmas and the upper unit consists of arc-type magmas (Puchtel et al., 1999). The authors suggested that the overthickened plume-related oceanic crust and the overlying arc-type magmas were later accreted to the existing TTG-type older crust of the Vodla block.

The ~2.8 Ka Kostomuksha greenstone belt in the NW Baltic Shield consists of komatiitic–basaltic submarine lavas and volcaniclastics whose formation was attributed by Puchtel et al. (1997) to plume activity and production of thick oceanic

crust–oceanic plateau. The plateau reached the Baltic continental margin but was too buoyant to subduct and became a new part of the continent. Using the example of the Cretaceous Caribbean plateau (Kerr et al., 1997), Puchtel and colleagues suggested that only the upper part of the thick komatiitic–basaltic crust of the proposed Baltic oceanic crust was imbricated and obducted on the ancient continent, whereas the deeper zones were subducted to the mantle. The subduction of this lower crust gave rise to subduction-related volcanism that erupted within the accreted sequences. This mechanism distinguishes the thick plateau-type oceanic crust from the thinner crust that is represented in the ophiolite sections. Moreover, it provides plausible explanation to the sequence of events that occurred within a 'cycle' of plateau formation–accretion–delamination–subduction–calc-alkaline arc-type magmatism.

9.07.6.6 Early (Hadean) Crust and the Initiation of Wilsonian Plate Tectonics–MOMO Cycles

The Nd and Hf isotope evolution trends (**Figures 11** and **12**) appear to deviate from the primitive (CHUR) mantle at the very early history of the Earth, calling for an early depletion event and possible formation of a complementary enriched reservoir.

The oldest zircons bearing evidence for early (Hadean) crust are the 4.4–4.1 Ga from Jack Hills, Australia (e.g., Peck et al., 2001). Other old zircons with ages lower than 4 Ga were found in the oldest remains of continental crust at Isua, Greenland (Nutman et al., 2009); Beartooth Mountain, the United States (Maier et al., 2012); and northern China (Cui et al., 2013; Geng et al., 2012). The interesting result that came from the Jack Hills zircons concerned their oxygen isotope values that are higher than the ~5–6‰ mantle value, indicating possible interaction with surface waters (Cavosie et al., 2005; Valley et al., 2002). This value indicates on a very early role of water at the Earth surface even before 4 Ga.

The series of works on the Hf and O isotopes in the Jack Hills and Mount Narryer in Western Australia reveal an exciting picture of mantle and crust evolution in the early Earth (e.g., Bell et al., 2011; Bell and Harrison, 2013; Blichert-Toft and Albarède, 2008; Harrison et al., 2005, 2008; Kemp et al., 2010, and see discussions on these results in Arndt, 2013). The zircons suggest a very early history of crustal formation on Earth that began by about 4.5 Ga (Bell et al., 2011). In the context of major themes discussed in this chapter review, they appear to lie on evolution trends that deviate from the MOMO evolution line (**Figure 12**). The oldest group appears to deviate at ~4.2 Ga and the younger one at ~3.7 Ga. The slope of the 'oldest ~4.2 array' (Lu/Hf ~ 0.02) is consistent with mafic composition. The slope of the 'younger ~3.7 Ga array' appears steeper and points to an enriched reservoir possibly similar to HREE-depleted TTG (Blichert-Toft and Albarède, 2008; Harrison et al., 2005, 2008). The slight deviation of the 'oldest array' could be consistent with the ^{142}Nd excess, indicating the existence of an early fractionation event and formation of an early mafic crust possibly during the solidification of a magma ocean (Boyet and Carlson, 2006).

Galer and Goldstein (1991) proposed that the early differentiation and depletion event was accompanied by the production of enriched alkali basalt crust (similar to that produced in intraplate rift-related volcanism) that had been gradually reincorporated into the mantle.

Boyet and Carlson (2006) noted that variable excesses of ^{142}Nd/^{144}Nd in rocks from Isua indicate on a depleted Hadean mantle and on mixing back of the enriched 'primordial crust' into the mantle shortly after 3.8 Ga, producing a ^{142}Nd/^{144}Nd homogeneous reservoir. This description is consistent with the clear separation between the two groups of Western Australian Hadean zircons. The apparent appearance of the ~3.7 Ga zircons on the MOMO evolution line is consistent with early derivation of continental crust from the MOMO mantle.

9.07.7 Summary

The continents come from the mantle. Large submarine oceanic plateaus whose production was associated with mantle plumes may have formed the basis for the growth of continental crust, at least since the late Archean time. The chronological evidence suggests that plateau formation was rapid and episodic and that it was followed by enhanced activity of arcs and calc-alkaline magmatism. Thick oceanic plateaus are buoyant and resist subduction. Yet some plateaus were subducted, reflecting probably internal differentiation within the plateaus and the geodynamic conditions (e.g., thermal state of the mantle). Subduction of segments of plateaus can in turn enhance arc magmatism, because its altered parts can increase the water flux at depth of magma generation. It should be also emphasized that oceanic plateaus may well be an important component of current arc magmatism, as some of the currently active arcs are located close to anomalously thick oceanic crust (e.g., the central Aleutians and Lesser Antilles).

Along with the production of new crust, the lithospheric mantle is evolved. It probably represents a frozen mantle wedge where the production of magmas and the migration of fluids and incompatible trace elements occur; thus, both the continental crust and the lithospheric mantle preserved the complicated history of continental growth. Besides the plateau and arc-accretion and calc-alkaline magmatism, delamination of lithospheric mantle and lower crust and underplating by rising asthenospheric magmas, as well as lithospheric melting, are all important processes that lead to the magmatic and thermal maturation of the continents. The fate of the continent relates eventually to the geodynamics and thermal conditions in the mantle.

Many of the previously mentioned topics require further study. In particular, we emphasize the importance of understanding the origin of the 'crustal type' V_p velocities that indicate a continental structure for some of the oceanic plateaus. Are these velocities associated with significant magmatic differentiation before accretion of the plateau to the existing continents? The mechanism of production of the calc-alkaline granitic batholiths is not as well established as is the composition of 'midcrust.' The processes of asthenospheric underplating, crustal foundering, and crustal recycling and their relation to the plate tectonic cycles and mantle dynamics through geologic time require more attention.

Finally, among the numerous works and thorough reviews on continental crust composition and evolution, we consider

and 'rephrase' the Taylor and Campbell statement 'no water, no oceans, no plate tectonics, no granites, no continents' as a major characteristic of the environment of production of the Earth continental crust.

References

Abbott D and Mooney W (1995) The structural and geochemical evolution of the continental-crust – Support for the oceanic plateau model of continental growth. *Reviews of Geophysics* 33: 231–242.

Abouchami W, Boher M, Michard A, and Albarède F (1990) A major 2.1 Ga event of Mafic magmatism in West Africa – an early stage of crustal accretion. *Journal of Geophysical Research-Solid Earth and Planets* 95: 17605–17629.

Albarède F (1998) The growth of continental crust. *Tectonophysics* 296: 1–14.

Albarède F and Van der Hilst RD (2002) Zoned mantle convection. *Philosophical Transactions of the Royal Society of London A* 360: 2569–2592.

Anderson AT (1982) Parental basalts in subduction zones; implications for continental evolution. *Journal of Geophysics Research* 87: 7047–7060.

Anderson DL (1994) Geology superplumes or supercontinents. *Geology* 22: 39–42.

Armstrong RL (1981a) Radiogenic isotopes – The case for crustal recycling on a near-steady-state no-continental-growth Earth. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 301: 443–472.

Armstrong RL (1981b) Comment on 'Crustal growth and mantle evolution: Inferences from models of element transport and Nd and Sr isotopes'. *Geochimica et Cosmochimica Acta* 45: 1251.

Armstrong RL (1988) Mesozoic and early Cenozoic magmatic evolution of the Canadian Cordillera. *Geological Society of America, Special Paper* 218: 55–91.

Armstrong RL (1991) The persistent myth of crustal growth. *Australian Journal of Earth Sciences* 38: 613–640.

Armstrong RL and Ward PL (1993) Late Triassic to early Eocene magmatism in North America Cordillera: Implications for the Western Interior Basin. In: Caldwell WGE and Kauffman EG (eds.) *Evolution of the Western Interior Basin*, *Geological Association of Canada Special Paper*, 39, pp. 49–72.

Arndt NT (2013) Formation and Evolution of the continental crust. In: Elliott T (ed.) *Geochemical, perspectives*, vol. 2, No 3. European Association of geochemistry, EAG.

Arth JG and Hanson GN (1975) Geochemistry and origin of early Precambrian crust of northeastern Minnesota. *Geochimica et Cosmochimica Acta* 39: 325–362.

Barth MG, McDonough WF, and Rudnick RL (2000) Tracking the budget of Nb and Ta in the continental crust. *Chemical Geology* 165: 197–213.

Beard JB and Lofgren GE (1989) Effect of water on the composition of partial melts of greenstone and amphibolites. *Science* 244: 195–197.

Bédard JH (2006) A catalytic delamination-driven model for coupled genesis of Archean crust and sub-continental lithospheric mantle. *Geochimica et Cosmochimica Acta* 70: 1188–1214.

Be'eri-Shlevin Y, Katzir Y, and Valley JW (2009) Crustal evolution and recycling in a juvenile continent: oxygen isotope ratio of zircon in the northern Arabian Nubian Shield. *Lithos* 107: 169–184.

Be'eri-Shlevin Y, Katzir Y, Blichert-Toft J, Kleinhanns IC, and Whitehouse MJ (2010) Nd-Sr-Hf-O isotope provinciality in the northernmost Arabian-Nubian Shield: Implications for crustal evolution. *Contributions to Mineralogy and Petrology* 160: 181–201.

Bell EH and Harrison TM (2013) Post-Hadean transitions in Jack Hills zircon provenance: A signal of the Late Heavy Bombardment? *Earth and Planetary Science Letters* 364: 1–11.

Bell EH, Harrison TM, McCulloch MT, and Young ED (2011) Early Archaean crustal evolution of the Jack Hills zircons. *Geochimica et Cosmochimica Acta* 75: 4816–4829.

Ben-Avraham Z, Nur A, Jones D, and Cox A (1981) Continental accretion and orogeny: From oceanic plateaus to allochthonous terrains. *Science* 213: 47–54.

Bentor YK (1985) The crustal evolution of the Arabian-Nubian Massif with special reference to the Sinai Peninsula. *Precambrian Research* 28: 1–74.

Bergantz GW (1989) Underplating and partial melting: Implications for melt generation and extraction. *Science* 245: 1093–1095.

Bickle MJ and Nisbet EG (1993) *The geology of the Belingwe Greenstone Belt, Zimbabwe*. Rotterdam: Balkema, pp 239.

Bielski M (1982) *Stages in the Evolution of the Arabian-Nubian Massif in Sinai*. PhD Thesis, Hebrew University, Jerusalem.

Bizzarro M, Simonetti A, Stevenson RK, and David J (2002) Hf isotope evidence for hidden mantle reservoir. *Geology* 30: 771–774.

Black R and Liegeois JP (1993) Cratons, mobile belts, alkaline rocks and continental lithosphere mantle: The Pan African testimony. *Journal of the Geological Society of London* 150: 89–98.

Blichert-Toft J and Albarède F (2008) Hafnium isotopes in Jack Hills zircons and the formation of the Hadean crust. *Earth and Planetary Science Letters* 265: 686–702.

Boher M, Abouchami W, Michard A, Albarède F, and Arndt NT (1992) Crustal growth in West Africa at 2.1 Ga. *Journal of Geophysical Research: Solid Earth* 97: 345–369.

Bohlen SR and Mezger K (1989) Origin of granulite terranes and the formation of the lowermost continental-crust. *Science* 244: 326–329.

Borg LE and Clynne MA (1998) The petrogenesis of felsic calc alkaline magmas from southernmost Cascades, California: Origin by partial melting of basaltic lower crust. *Journal of Petrology* 39: 1197–1222.

Bowring SA, Housh TB, and Isachsen CE (1990) The Acasta gneisses: Remnants of earth earliest crust. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 319–343. New York: Oxford University Press.

Boyet M and Carlson RW (2005) ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.

Boyet M and Carlson RW (2006) A new geochemical model for the Earth's mantle inferred from 146Sm-142Nd systematics. *Earth and Planetary Science Letters* 250: 254–268.

Brown GC (1979) The changing patterns of batholith emplacement during Earth history. In: Atherton MP (ed.) *Origin of Granitic Batholiths, Geochemical Evidence*, pp. 106–115. UK: Shiva.

Campbell IH and Taylor SR (1983) No Water, No Granites – No Oceans, No Continents. *Geophysical Research Letters* 10: 1061–1064.

Card KD (1990) A review of the Superior province of the Canadian Shield, a product of Archean accretion. *Precambrian Research* 48: 99–156.

Carlson RL, Christensen NI, and Moore RP (1980) Anomalous crustal structures on ocean basins: Continental fragments and oceanic plateaus. *Earth and Planetary Science Letters* 51: 171–180.

Cavosie AJ, Valley JW, and Wilde SA (2005) Magmatic 18O in 4400–3900 Ma detrital zircons: A record of the alteration and recycling of crust in the Early Archean. *Earth and Planetary Science Letters* 235: 663–668.

Chappell BW, White AJR, and Wyborn D (1987) The importance of residual source material (restite) in granite petrogenesis. *Journal of Petrology* 28: 1111–1138.

Clarke FW (1889) The relative abundance of the chemical elements. *Philosophical Society of Washington Bulletin* XI: 131–142.

Clarke FW and Washington HS (1924) The composition of the Earth's crust. *USGS Professional Paper* 127: 117.

Clemens JD (1989) Granulites still being formed. *Nature* 342: 228–229.

Cloos M (1993) Lithospheric buoyancy and collisional orogenesis – Subduction of Oceanic Plateaus, Continental Margins, Island Arcs, Spreading Ridges, and Seamounts. *Geological Society of America Bulletin* 105: 715–737.

Clowes RM, Hammer PTC, Fernández-Viejo G, and Welford JK (2005) Lithospheric structure in northwestern Canada from Lithoprobe seismic refraction and related studies: A synthesis. *Canadian Journal of Earth Sciences* 42: 1277–1293.

Collins WJ, Beams SD, White AJR, and Chappell BW (1982) Nature and origin of A-type granites with particular reference to southeastern Australia. *Contributions to Mineralogy and Petrology* 80: 189–200.

Colpron M, Nelson J-AL, and Murphy DC (2007) Northern Cordilleran terrains and their interactions through time. *GSA Today* 17(4/5): 4–10. http://dx.doi.org/10.1130/GSAT01704-5A.1.

Condie KC (1990) Growth and accretion of continental crust: Inferences based on Laurentia. *Chemical Geology* 83: 183–194.

Condie KC (1993) Chemical-composition and evolution of the upper continental crust – Contrasting results from surface samples and shales. *Chemical Geology* 104: 1–37.

Condie KC and Aster RC (2010) Episodic zircon age spectra of orogenic granitoids: The supercontinent connection of continental growth. *Precambrian Research* 180: 227–236.

Condie KC, Belousova EA, Griffin WL, and Sircombe KN (2009) Granitoid events in space and time: Constraints from igneous and detrital zircon age spectra. *Gondwana Research* 15: 228–242.

Condie KC, Bickford ME, Aster RC, Belousova EA, and Scholl DW (2011) Episodic zircon ages, Hf isotope composition, and the preservation rate of continental crust. *Geological Society of America Bulletin* 123: 951–957.

Coney PJ, Jones DL, and Monger JWH (1980) Cordilleran suspect terranes. *Nature* 288: 329–333.

Cook FA and Erdmer P (2005) An 1800 km cross section of the lithosphere through the northwestern North American plate: Lessons from 4.0 billion years of Earth's history. *Canadian Journal of Earth Sciences* 42: 1295–1311.

Costa S and Rey P (1995) Lower crustal rejuvenation and growth during post-thickening collapse; insights from a crustal cross section through a Variscan metamorphic core complex. *Geology* 23: 905–908.

Creaser RA, Price RC, and Wormald RJ (1991) A-type granites revisited: Assessment of a residual-source model. *Geology* 19: 163–166.
Cross TA and Pilger RH (1978) Constraints on absolute motion and plate interaction inferred from Cenozoic igneous activity in western United-States. *American Journal of Science* 278: 865–902.
Cui PL, Sun JG, Sha D, Wang XJ, Zhang O, Gu AL, and Wang ZY (2013) Oldest zircon xenocryst (4.17 Ga) from the North China Craton. *International Geology Review* 55: 1902–1908. http://dx.doi.org/10.1080/00206814.2013.805925.
De Wit MJ, Hart RA, and Hart RJ (1987) The Jamestown ophiolite complex Barberton mountain belt: A section through 3.5 Ga oceanic crust. *Journal of African Earth Science* 6: 681–730.
DePaolo DJ (1981) Neodymium isotopes in the Colorado Front Range and crust-mantle evolution in the Proterozoic. *Nature* 291: 193–196.
DePaolo DJ, Linn AM, and Schubert G (1991) The continental crustal age distribution – Methods of determining mantle separation ages from Sm-Nd isotopic data and application to the southwestern United-States. *Journal of Geophysical Research* 96: 2071–2088.
DePaolo DJ and Wasserburg GJ (1976) Nd isotopic variations and petrogenetic models. *Geophysical Research Letters* 3: 249–252.
Dewey JF and Windley BF (1981) Growth and Differentiation of the Continental Crust. *Philosophical Transactions of the Royal Society of London* 301: 189–206.
Dhuime B, Hawkesworth CJ, and Cawood P (2011) When continents formed. *Science* 331: 154–155.
Droop GTR, Clemens JD, and Dalrymple DJ (2003) Processes and conditions during contact anatexis, melt escape and restite formation: The Huntley Gabbro complex, NE Scotland. *Journal of Petrology* 44: 995–1029.
Drummond MS, Defant MJ, and Kepenzhinskas PK (1996) Petrogenesis of slab-derived Trondhjemite-tonalite-dacite/adakite magmas. *Transactions of the Royal Society of Edinburgh Earth Science* 87: 205–205.
Eby GN (1990) The A-type granitoids: A review of their occurrence and chemical characteristics and speculations on their petrogenesis. *Lithos* 26: 115–134.
Eldholm O and Coffin MF (2000) Large igneous provinces and plate tectonics. *AGU Monograph* 121: 309–326.
Eyal M, Litvinovsky B, Jahn BM, Zanvilevich A, and Katzir Y (2010) Origin and evolution of post-collisional magmatism: Coeval Neoproterozoic calc-alkaline and alkaline suites of the Sinai Peninsula. *Chemical Geology* 269: 153–179.
Eyal M, Litvinovsky BA, Katzir Y, and Zanvilevich AN (2004) The Pan-African high-K calc-alkaline peraluminous Elat granite from southern Israel: Geology, geochemistry and petrogenesis. *Journal of African Earth Sciences* 40: 115–136.
Galer SJG (1991) Interrelationships between continental freeboard, tectonics and mantle temperature. *Earth and Planetary Science Letters* 105: 214–228.
Galer SJG and Goldstein SL (1991) Early mantle differentiation and its thermal consequences. *Geochimica et Cosmochimica Acta* 55: 227–239.
Gallet S, Jahn BM, Lanoe BV, Dia A, and Rossello E (1998) Loess geochemistry and its implications for particle origin and composition of the upper continental crust. *Earth and Planetary Science Letters* 156: 157–172.
Gao S, Luo TC, Zhang BR, et al. (1998) Chemical composition of the continental crust as revealed by studies in East China. *Geochimica et Cosmochimica Acta* 62: 1959–1975.
Geng Y, Du L, and Ren L (2012) Growth and reworking of the early Precambrian continental crust in the North China Craton: Constraints from zircon Hf isotopes. *Gondwana Research* 21: 517–529.
Goldschmidt VM (1933) Grundlagen der quantitativen Geochemie. *Fortschr. Mienral. Kirst. Petrogr* 17: 112.
Goldschmidt VM (1958) *Geochemistry*. Oxford: Oxford University Press.
Goodwin AM (1991) *Precambrian Geology*. London: Academic Press.
Green TH and Ringwood AE (1968) Genesis of the calc-alkaline igneous rock suite. *Contribution to Mineralogy and Petrology* 18: 105–162.
Green MG, Sylvester PJ, and Buick R (2000) Growth and recycling of early Archaean continental crust: Geochemical evidence from the Coonterunah and Warrawoona Groups, Pilbara craton, Australia. *Tectonophysics* 322: 69–88.
Griffin WL and O'Reilly SY (1987) Is the continental Moho the crust-mantle boundary. *Geology* 15: 241–244.
Grove TL, Chatterjee N, Parman SW, and Medard E (2006) The influence of H2O on mantle wedge melting. *Earth and Planetary Science Letters* 249: 74–89.
Harrison TM, Müller W, Blichert-Toft J, Albarède F, Holden P, and Mojzsis SJ (2005) Heterogeneous Hadean hafnium: Evidence of continental crust by 4.4-4.5 Ga. *Science* 310: 1947–1950.
Harrison TM, Schmitt AK, McCulloch MT, and Lovera OM (2008) Early ($\geq$4.5 Ga) formation of terrestrial crust Lu-Hf, 18O and Ti thermometry results for Hadean zircons. *Earth and Planetary Science Letters* 268: 476–486.
Hart SR, Hauri EH, Oschmann LA, and Whitehead JA (1992) Mantle plumes and entrainment: The isotopic evidence. *Science* 256: 517–520.
Hattori Y, Suzuki K, Honda M, and Shimizu H (2003) Re-Os isotope systematics of the Taklimakan Desert sands, moraines and river sediments around the Taklimakan Desert, and of Tibetan soils. *Geochimica et Cosmochimica Acta* 67: 1195–1205.
Hawkesworth CJ, Dhuime B, Pietranik AB, Cawood P, Kemp AIS, and Storey CD (2010) The generation and evolution of the continental crust. *Journal of the Geological Society of London* 167: 229–248.
Hawkesworth CJ, Erlank AJ, Marsh JS, Menzies MA, and van Calsteren P (1983) Evolution of the continental lithosphere: Evidence from volcanics and xenoliths in southern Africa. In: Hawkesworth CJ and Norry MJ (eds.) *Continental Basalts and Mantle Xenoliths*, pp. 111–138. Nantwich: Shiva.
Hawkesworth CJ and Kemp AIS (2006) Evolution of the continental crust. *Nature* 443: 811–817.
Hawkesworth CJ, Cawood PA, Kemp AIS, Storey C, and Dhuime B (2009) A matter of preservation. *Science* 323: 49–50.
Helz RT (1976) Phase relations of basalts in their melting ranges at PH-20 = 5 kB. 2. Melt compositions. *Journal of Petrology* 17: 139–193.
Hofmann AW (1988) Chemical differentiation of the Earth: The relationship between mantle, continental crust, and oceanic crust. *Earth and Planetary Science Letters* 90: 297–314.
Hofmann AW, Jochum KP, Seufert M, and White WM (1986) Nb and Pb in oceanic basalts: New constraints on mantle evolution. *Earth and Planetary Science Letters* 79: 33–45.
Hollister LS and Andronicos CL (2006) Formation of new continental crust in Western British Columbia during transpression and transtension. *Earth and Planetary Science Letters* 249: 29–38.
Holloway JR and Burnham CW (1972) Melting relations of basalt with equilibrium water pressure less than total pressure. *Journal of Petrology* 13: 1–29.
Huppert HE and Sparks RSJ (1988) The generation of granitic magmas by intrusion of basalts into continental crust. *Journal of Petrology* 29: 259–275.
Hurley PM and Rand JR (1969) Pre-drift continental nuclei. *Science* 164: 1229–1242.
Iizuka T, Campbell IH, Allen CM, Gill JB, Maruyama S, and Makoka F (2013) Evolution of the African continental crust as recorded by U-Pb, Lu-Hf and O isotopes in detrital zircons from modern rivers. *Geochimica et Cosmochimica Acta* 107: 96–120.
Ishikawa A, Shigenori M, and Komiya T (2004) layered lithospheric mantle beneath the Ontong Java plateau: Implications from xenoliths in Alonite, Malaita, Solomon Islands. *Journal of Petrology* 47: 2011–2044.
Jacobsen SB and Wasserburg GJ (1980) Sm-Nd isotopic evolution of chondrites. *Earth and Planetary Science Letters* 50: 139–155.
Johnston ST and Borel GD (2007) The odyssey of the Cache Creek terrain, Canadian Cordillera: Implications for accretionary orogens, tectonic setting of Panthalassa, the Pacific superswell, and break-up of Pangea. *Earth and Planetary Science Letters* 253: 415–428.
Kay RW and Kay SM (1981) The nature of the lower continental-crust – Inferences from geophysics, surface geology, and crustal xenoliths. *Reviews of Geophysics* 19: 271–297.
Kelemen PB (1995) Genesis of high Mg-number Andesites and the continental-crust. *Contributions to Mineralogy and Petrology* 120: 1–19.
Kemp AIS, Hawkesworth CJ, Paterson BA, and Kinny PD (2006) Episodic growth of the Gondwana supercontinent from hafnium and oxygen isotopes in zircon. *Nature* 439: 580–583.
Kemp AIS, Wilde SA, Hawkesworth CJ, et al. (2010) Hadean crustal evolution revisited: New constraints from Pb-Hf isotope systematic of the Jack Hills zircons. *Earth and Planetary Science Letters* 296: 45–56.
Kerr AC (2013) Oceanic plateaus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, 2nd edn., pp. 171–275. Amsterdam: Elsevier.
Kerr AC, Tarney J, Marriner GF, Nivia A, and Saunders AD (1997) The Caribbean-Colombian Cretaceous igneous province: The internal anatomy of an oceanic plateau. *Large Igneous Provinces; Continental, Oceanic and Planetary Flood Volcanism*, *American Geophysical Union Monograph*, 100, pp. 45–93.
Kessel R, Stein M, and Navon O (1998) Petrogenesis of late Neoproterozoic dikes in the northern Arabian-Nubian Shield – Implications for the origin of A-type granites. *Precambrian Research* 92: 195–213.
Klemperer SL, Hauge TA, Hauser EC, Oliver JE, and Potter CJ (1986) The Moho in the northern Basin and Range province along the COCORP 40°N seismic reflection transect. *Geological Society of America Bulletin* 97: 603–618.
Kroenke LW (1974) Origin of continents through development and coalescence of oceanic flood basalt plateaus. *Transactions American Geophysical Union* 55: 443.
Langmuir CH, Klein EM, and Plank T (1992) Petrological constraints on mid-ocean ridge basalts. In: Mantle flow and melt generation at mid-ocean ridges, *Geophysical Monograph* 71: 182–280.
Larson RL (1991) Latest pulse of Earth – Evidence for a Mid-cretaceous superplume. *Geology* 19: 547–550.

Lassiter JC, Depaolo DJ, and Mahoney JJ (1995) Geochemistry of the Wrangellia flood-basalt province – Implications for the role of continental and oceanic lithosphere in flood-basalt genesis. *Journal of Petrology* 36: 983–1009.
Mahoney JJ, Storey M, Duncan RA, Spencer KJ, and Pringle M (1993a) Geochemistry and age of the Ontong Java Plateau. In: The Mesozoic Pacific: Geology, Tectonics, and Volcanism, *American Geophysical Union Monograph*, 77: 233–261.
Mahoney JJ, Storey M, Duncan RA, Spencer KJ, and Pringle M (1993b) Geochemistry and geochronology of Leg 130 basement lavas; nature and origin of the Ontong Java Plateau. *Proceedings of the Ocean Drilling Program, Scientific Results* 130: 3–22.
Maier AC, Cates NL, Trail D, and Mojzsis SJ (2012) Geology, age and field relations of Hadean-zircon bearing supracrustal rocks from Quad Creek, eastern Beartooth Mountain (Montana and Wyoming, USA). *Chemical Geology* 312: 47–57.
Mason WG, Moresi L, Betts PG, and Miller MS (2010) Three-dimensional numerical models of the influence of a buoyant oceanic plateau on subduction zones. *Tectonophysics* 483: 71–79.
McCulloch MT and Bennett VC (1994) Progressive growth of the Earths continental-crust and depleted mantle – geochemical constraints. *Geochimica et Cosmochimica Acta* 58: 4717–4738.
McCulloch MT and Wasserburg GJ (1978) Sm-Nd and Rb-Sr chronology of continental crust formation. *Science* 200: 1003–1011.
McGeary S, Nur A, and Ben-Avraham Z (1985) Spatial gaps in arc volcanism – the effect of collision or subduction of oceanic plateaus. *Tectonophysics* 119: 195–221.
McKenzie D and Bickle MJ (1988) The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology* 29: 625–679.
McKenzie D and O'Nions RK (1983) Mantle reservoirs and ocean island basalts. *Nature* 301: 229–231.
McLennan SM (1988) Recycling of the continental-crust. *Pure and Applied Geophysics* 128: 683–724.
Mezger K (1992) Temporal evolution of regional granulite terrains: Implications for the evolution of the lowermost crust. In: Fountain DM, Arculus R, and Kay RV (eds.) *Continental Lower Crust*, pp. 447–478. New York: Elsevier Science.
Monger JWH (1993) Canadian cordilleran tectonics: From geosynclines to crustal collage. *Canadian Journal of Earth Sciences* 30: 209–231.
Moorbath S (1975) Evolution of Precambrian crust from strontium isotopic evidence. *Nature* 254: 395–398.
Moorbath S (1978) Age and isotope evidence for evolution of continental crust. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 288: 401–412.
Morag N, Avigad D, Gerdes A, Belousova E, and Harlavan Y (2011) Crustal evolution and recycling in the northern Arabian-Nubian Shield: New perspectives from zircon Lu-Hf and U-Pb systematics. *Precambrian Research* 186: 101–116.
Moyen JE and Martin H (2012) Forty years of TTG research. *Lithos* 148: 312–336.
Mushkin A, Navon O, Halicz L, Hartmann G, and Stein M (2003) The petrogenesis of A-type magmas from the Amram Massif, southern Israel. *Journal of Petrology* 44: 815–832.
Myers JS and Swagers C (1997) The Yilgarn Craton. In: de Wit MJ and Ashwal LD (eds.) *Greenstone Belts*, pp. 640–656. Oxford: Oxford Science Publication.
Neal CR, Mahoney JJ, Kroenke LW, Duncan RA, and Petterson MG (1997) The Ontong Java Plateau. In: Mahoney JJ and Coffin M (eds.) *On Large Igneous Provinces*, *American Geophysical Union Monograph*, 100, pp. 83–216.
Nelson BK and DePaolo DJ (1985) Rapid production of continental-crust 1.7 to 1.9 b.y. ago – Nd isotopic evidence from the basement of the North-American mid-continent. *Geological Society of America Bulletin* 96: 746–754.
Nur A and Ben-Avraham Z (1977) Lost Pacifica continent. *Nature* 270: 41–43.
Nur A and Ben-Avraham Z (1978) Speculations on mountain building and the lost Pacifica continent. *Journal of Physics of the Earth Supplements* 26: S21–S37.
Nur A and Ben-Avraham Z (1981) Volcanic gaps and the consumption of aseismic ridges in South America. *Memoir – Geological Society of America* 154: 729–740.
Nur A and Ben-Avraham Z (1982) Displaced terranes and mountain building. In: Hsu KJ (ed.) *Mountain Building Processes*, pp. 73–84. London: Academic Press.
Nutman AP, Friend CRL, and Paxton S (2009) Detrital zircon sedimentary provenance ages for the Eoarchaean Isua supercrustal belt southern West Greenland: Juxtaposition of an imbricated ca, 3700 Juvenile arc against an older complex with 3920–3760 Ma components. *Precambrian Research* 172: 212–233.
Nye CJ and Reid MR (1986) Geochemistry of primary and least fractionated lavas from Okmok Volcano, Central Aleutians – Implications for arc magmagenesis. *Journal of Geophysical Research-Solid Earth and Planets* 91: 271–287.
O'Nions RK, Evensen NM, and Hamilton PJ (1979) Geochemical modeling of mantle differentiation and crust growth. *Journal of Geophysical Research* 84: 6091–6101.
Oxburgh ER and Parmantier EM (1977) Compositional density stratification in the oceanic lithosphere-causes and consequences. *Journal of Geological Society of London* 133: 343–355.
Patchett PJ (1996) Crustal growth – Scum of the Earth after all. *Nature* 382: 758–759.
Patchett PJ and Arndt N (1986) Nd isotopes and tectonics of 1.9–1.7 Ga crustal genesis. *Earth and Planetary Science Letters* 78: 329–338.
Patchett PJ and Chase CG (2002) Role of transform continental margins in major crustal growth episodes. *Geology* 30: 39–42.
Patchett PJ and Samson SD (2003) Ages and Growth of the Continental Crust from Radiogenic Isotopes. In: Holland HD and Turekian (eds.) *Treatise on Geochemistry*, vol. 3, pp. 321–348. Oxford: Elsevier-Pergamon.
Patchett JP, Vervoort JD, and Söderlund U (2004) Lu-Hf and Sm-Nd isotopic systematics in chondrites and their constraints on the Lu-Hf properties of the earth. *Earth and Planetary Science Letters* 222: 29–41.
Patchett PJ, White WM, Feldmann H, Kielinczuk S, and Hofmann AW (1984) Hafnium rare-Earth element fractionation in the sedimentary system and crustal recycling into the Earth's mantle. *Earth and Planetary Science Letters* 69: 365–378.
Pearcy LG, Debari SM, and Sleep NH (1990) Mass balance calculations for two sections of island arc crust and implications for the formation of continents. *Earth and Planetary Science Letters* 96: 427–442.
Peck WH, Valley JW, Wilde SA, and Graham CM (2001) Oxygen isotope ratios and rare earth elements in 3.3 to 4.4 Ga zircons: Ion microprobe evidence for the high δ18O continental crust and oceans in the early Archean. *Geochimica et Cosmochimica Acta* 65: 4215–4229.
Percival JA, Stern RA, Skulski T, Card KD, Mortensen JK, and Bégin NJ (1994) Minto block, Superior province: Missing link in deciphering assembly of the craton at 2.7 Ga. *Geology* 22: 839–842.
Pilger RH (1978) Method for finite plate reconstructions, with applications to Pacific-Nazca plate evolution. *Geophysical Research Letters* 5: 469–472.
Plank T (2005) Constraints from thorium/lanthanum on sediment recycling at subduction zones and the evolution of the continents. *Journal of Petrology* 46: 921–944.
Plank T and Langmuir CH (1998) The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chemical Geology* 145: 325–394.
Polat A and Kerrich R (2001a) Geodynamic processes, continental growth, and mantle evolution recorded in late Archean greenstone belts of the southern Superior Province, Canada. *Precambrian Research* 112: 5–25.
Polat A and Kerrich R (2001b) Magnesian andesites, Nb-enriched basalt-andesites, and adakites from late-Archean 2.7 Ga Wawa greenstone belts, Superior Province, Canada: Implications for late Archean subduction zone petrogenetic processes. *Contributions to Mineralogy and Petrology* 141: 36–52.
Polat A, Kerrich R, and Wyman DA (1998) The late Archean Schreiber-Hemlo and White River Dayohessarah greenstone belts, Superior Province: Collages of oceanic plateaus, oceanic arcs, and subduction-accretion complexes. *Tectonophysics* 289: 295–326.
Puchtel IS, Arndt NT, Hofmann AW, et al. (1998a) Petrology of mafic lavas within the Onega plateau, central Karelia: Evidence for 2.0 Ga plume-related continental crustal growth in the Baltic Shield. *Contributions to Mineralogy and Petrology* 130: 134–153.
Puchtel IS, Haase KM, Hofmann AW, et al. (1997) Petrology and geochemistry of crustally contaminated komatiitic basalts from the Vetreny Belt, southeastern Baltic Shield: Evidence for an early Proterozoic mantle plume beneath rifted Archean continental lithosphere. *Geochimica et Cosmochimica Acta* 61: 1205–1222.
Puchtel IS, Hofmann AW, Amelin YV, Garbe-Schonberg CD, Samsonov AV, and Schipansky AA (1999) Combined mantle plume-island arc model for the formation of the 2.9 Ga Sumozero-Kenozero greenstone belt, SE Baltic Shield: Isotope and trace element constraints. *Geochimica et Cosmochimica Acta* 63: 3579–3595.
Puchtel IS, Hofmann AW, Mezger K, Jochum KP, Shchipansky AA, and Samsonov AV (1998b) Oceanic plateau model for continental crustal growth in the archaean, a case study from the Kostomuksha greenstone belt, NW Baltic Shield. *Earth and Planetary Science Letters* 155: 57–74.
Puchtel IS, Zhuravlev DZ, Samsonov AV, and Arndt NT (1993) Petrology and geochemistry of metamorphosed komatiites and basalts from the Tungurcha greenstone belt, Aldan Shield. *Precambrian Research* 62: 399–418.
Reischmann T, Kroner A, and Basahel A (1983) Petrography, geochemistry and tectonic setting of metavolcanic sequences from the Al Lith area, southwestern Arabian shield. *Bulletin of the Faculty of Earth Sciences King Abdulaziz University* 6: 365–378.
Rey P (1993) Seismic and tectonometamorphic characters of the lower continental crust in Phanerozoic areas: A consequence of post-thickening extension. *Tectonics* 12: 580–590.
Reymer A and Schubert G (1984) Phanerozoic addition rates to the continental-crust and crustal growth. *Tectonics* 3: 63–77.
Reymer A and Schubert G (1986) Rapid growth of some major segments of continental crust. *Geology* 14: 299–302.
Richards MA, Duncan RA, and Courtillot VE (1989) Flood basalts and hot-spot tracks – Plume heads and tails. *Science* 246: 103–107.

Richardson WP, Emile AB, Okal EA, and Van der Le S (2000) Rayleigh-wave tomography of the Ontong-Java Plateau. *Physics of the Earth and Planetary Interiors* 118: 29–51.
Rocholl A, Stein M, Molzahn M, Hart SR, and Worner G (1995) Geochemical evolution of rift magmas by progressive tapping of stratified mantle source beneath the Ross Sea Rift, Antarctica. *Earth and Planetary Science Letters* 131: 207–224.
Rudnick RL (1992) Xenoliths-samples of the lower continental crust. In: Fountain DM, Arculus R, and Kay RW (eds.) *Continental Lower Crust*, pp. 269–316. Amsterdam: Elsevier.
Rudnick RL (1995) Making continental-crust. *Nature* 378: 571–578.
Rudnick RL and Fountain DM (1995) Nature and composition of the continental crust: A lower crustal perspective. *Reviews of Geophysics* 33: 267–310.
Rudnick RL and Gao S (2003) *Composition of the Continental Crust*. *Treatise on Geochemistry*. Elsevier.
Rudnick RL and Taylor R (1987) The composition and petrogenesis of the lower continental crust: A xenolith study. *Journal of Geophysical Research* 92: 13981–14005.
Rushmer T (1991) Partial melting of two amphibolites: Contrasting experimental results under fluid-absent conditions. *Journal of Geophysical Research* 92: 13981–14055.
Samson SD, Mcclelland WC, Patchett PJ, Gehrels GE, and Anderson RG (1989) Evidence from neodymium isotopes for mantle contributions to Phanerozoic crustal genesis in the Canadian Cordillera. *Nature* 337: 705–709.
Samson SD, Patchett PJ, Gehrels GE, and Anderson RG (1990) Nd and Sr Isotopic characterization of the Wrangellia Terrane and implications for crustal growth of the Canadian Cordillera. *The Journal of Geology* 98: 749–762.
Samson SD, Patchett PJ, Mcclelland WC, and Gehrels GE (1991) Nd isotopic characterization of metamorphic rocks in the Coast Mountains, Alaskan and Canadian Cordillera: Ancient crust bounded by juvenile terranes. *Tectonics* 10: 770–780.
Schubert G and Reymer A (1985) Continental volume and freeboard through geological time. *Nature* 316: 336–339.
Schubert G and Sandwell D (1989) Crustal volumes of the continents and of oceanic and continental submarine plateaus. *Earth and Planetary Science Letters* 92: 234–246.
Shirey SB and Hanson GN (1984) Mantle-derived Archean monozodiorites and trachyandesites. *Nature* 310: 222–224.
Sleep NH and Windley BF (1982) Archean plate tectonic constraints and inferences. *Journal of Geology* 90: 363–379.
Smith HS and Erlank AJ (1982) Geochemistry and petrogenesis of komatiites from the Barberton greenstone belt. In: Arndt NT and Nisbet EG (eds.) *Komatiites*, pp. 347–398. Allen and Unwin.
Smithies RH, Howard HM, Evins OM, et al. (2011) High-temperature granite magmatism, crust-mantle interaction and the mesoproterozoic intracontinental evolution of the Musgrave Province, Central Australia. *Journal of Petrology* 52: 931–958.
Sparks RSJ, Sigurdsson H, and Wilson L (1977) Magma mixing: a mechanism for triggering acid explosive eruptions. *Nature* 267: 315–318.
Stein M (2003) Tracing the plume material in the Arabian-Nubian Shield. *Precambrian Research* 123: 223–234.
Stein M (2006) The rising of the asthenospheric MORB mantle into the rifting Arabian plate and its consequences: Heating, melting uplifting. In: *IAVCEI Conference Guangzhou* .
Stein M and Goldstein SL (1996) From plume head to continental lithosphere in the Arabian-Nubian shield. *Nature* 382: 773–778.
Stein M and Hofmann AW (1992) Fossil plumes beneath the Arabian lithosphere? *Earth and Planetary Science Letters* 114: 193–209.
Stein M and Hofmann AW (1994) Mantle plumes and episodic crustal growth. *Nature* 372: 63–68.
Stein M, Navon O, and Kessel R (1997) Chromatographic metasomatism of the Arabian-Nubian lithosphere. *Earth and Planetary Science Letters* 152: 75–91.
Stern RJ (1981) Petrogenesis and tectonic setting of late Precambrian ensimatic volcanic rocks. Central Eastern Desert of Egypt. *Precambrian Research* 16: 195–230.
Stern RJ and Gottfried D (1986) Petrogenesis of a Late Precambrian (575–600 Ma) bimodal suite in Northeast Africa. *Contributions to Mineralogy and Petrology* 92: 492–501.
Stern CR and Kilian R (1996) Role of the subducted slab, mantle wedge and continental crust in the generation of adakites from the Andean Austral Volcanic Zone. *Contributions to Mineralogy and Petrology* 123: 263–281.
Sylvester PJ, Campbell IH, and Bowyer DA (1997) Niobium/uranium evidence for early formation of the continental crust. *Science* 275: 521–523.
Taylor SR (1967) The origin and growth of continents. *Tectonophysics* 4: 17–34.
Taylor SR and McLennan SM (1985) *The Continental Crust: Its Composition and its Evolution*. Oxford: Blackwell.
Taylor SR and Mclennan SM (1995) The geochemical evolution of the continental-crust. *Reviews of Geophysics* 33: 241–265.
Taylor SR and McLennan SM (1996) The evolution of continental crust. *Scientific American* 274: 76–81.
Tera F, Brown L, Morris J, Selwayn Sacks I, Klein J, and Middleton R (1986) Sediment incorporation in island-arc magmas: Inferences from 10Be. *Geochimica et Cosmochimica Acta* 50: 535–550.
Tetreault JL and Buiter SJH (2012) Geodynamic models of terrain accretion: Testing the fate of island arcs, oceanic plateaus, and continental fragments in subduction zones. *Journal of Geophysics Research* 117: B08403. http://dx.doi.org/10.1029/2012JB009316.
Turcotte DL and Schubert G (1982) *Geodynamics*. New York: Wiley.
Turner SP, Foden JD, and Morrison RS (1992) Derivation of some A-type magmas by fractionation of basaltic magma: An example from the Padthaway Ridge, South Australia. *Lithos* 28: 151–179.
Valley JM, Peck WH, King EM, and Wilde SA (2002) A cool early Earth. *Geology* 30: 351–354.
van Keken PE, Hauri EH, and Ballentine CJ (2002) Mantle mixing: The generation, preservation, and destruction of chemical heterogeneity. *Annual Reviews of Earth and Planetary Science Letters* 30: 493–525.
Van Kranendonk MJ (2010) Two types of Archean continental crust: Plume and plate tectonics on early Earth. *American Journal of Science* 310: 1187–1209.
Volkert RA, Feigenson MD, Patino LC, Delaney JC, and Drake AA (2000) Sr and Nd isotopic compositions, age and petrogenesis of A-type granitoids of the Vernon Supersuite, New Jersey Highlands, USA. *Lithos* 50: 325–347.
Weinstein Y, Navon O, Altherr R, and Stein M (2006) The role of lithospheric mantle heterogeneity in the generation of Plio-Pleistocene alkali basalt suites from NW Harrat Ash Shaam (Israel). *Journal of Petrology* 47: 1017–1050.
Weismann A, Kessel R, Navon O, and Stein M (2013) The petrogenesis of calc-alkaline granites from the Elat massif, Northern Arabian Nubian Shield. *Precambrian Research* 236: 252–264.
Whalen JB, Currie KL, and Chappell BW (1987) A-type granites: geochemical characteristics, discrimination and petrogenesis. *Contributions to Mineralogy and Petrology* 95: 407–419.
White AJR and Chappell BW (1977) Ultrametamorphism and granitoid genesis. *Tectonophysics* 43: 7–22.
White WM and Patchett PJ (1984) Hf-Nd-Sr isotopes and incompatible-element abundances in island arcs; implications for magma origins and crust- mantle evolution. *Earth and Planetary Science Letters* 67: 167–185.
White WM, Dupre B, and Vidal P (1985) Isotope and trace element geochemistry of sediments from the Barbados Ridge-Demerara Plain region, Atlantic Ocean. *Geochimica et Cosmochimica Acta* 49: 1875–1886.
Wilde SA, Valley JW, Peck WH, and Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409: 175–178.
Windley BF (1995) *The Evolving Continents*. Chichester: Wiley.
Wise DU (1974) Continental margins, freeboard and volumes of continents and oceans through time. In: Burke CA and Drake CL (eds.) *The Geology of Continental Margins*, pp. 45–58. Berlin: Springer.
Wolf MB and Wyllie PJ (1994) Dehydration of amphibolite at 10 kbar: the effects of temperature and time. *Contributions to Mineralogy and Petrology* 115: 369–383.
Wörner G, Zindler A, Staudigel H, and Schmincke HU (1986) Sr, Nd, and Pb isotope geochemistry of Tertiary and Quaternary alkaline volcanics from West Germany. *Earth and Planetary Science Letters* 79: 107–119.
Wyborn D and Chappell BW (1986) The petrogenetic significance of chemically related plutonic and volcanic rock units. *Geological Magazine* 123: 619–628.
Wyllie PJ (1977) Crustal anatexis: An experimental review. *Tectonophysics* 43: 41–71.
Wyllie PJ (1984) Sources of granitoid magmas at convergent plate boundaries. *Physics of the Earth and Planetary Interiors* 35: 12–18.
Wyllie PJ, Huang W, Stern CR, and Maaloe S (1976) Granitic magmas: Possible and impossible sources, water contents, and crystallization sequences. *Canadian Journal of Earth Sciences* 13: 1007–1019.
Zimmer 329, Kroner A, and Jochum KP (1995) The Gabal Gerf complex, Nubian Shield, northeast Africa: Geochemistry, Petrology, Isotope systematic and tectonic setting of a large, late Proterozoic, N-MORB-like ophiolite. *Chemical Geology* 123: 29–51.
Zorpi MJ, Coulon C, Orsini JB, and Cocirta C (1989) Magma mingling, zoning and emplacement in calc alkaline granitoid plutons. *Tectonophysics* 157: 315–329.

This page intentionally left blank

9.08 Thermal and Compositional Evolution of the Core

F Nimmo, University of California, Santa Cruz, CA, USA

© 2015 Elsevier B.V. All rights reserved.

9.08.1	**Introduction**	201
9.08.2	**Present-Day State of the Core**	202
9.08.2.1	Density and Pressure	202
9.08.2.2	Thermodynamic Properties	202
9.08.2.3	Composition	202
9.08.2.3.1	Light elements	203
9.08.2.3.2	Isotopes	203
9.08.2.4	Temperature Structure	204
9.08.2.5	Inner-Core Structure	205
9.08.2.6	The CMB Region	205
9.08.2.7	Dynamo Behavior Over Time	206
9.08.2.7.1	Ohmic dissipation	206
9.08.2.8	Summary	207
9.08.3	**Evolution of the Core**	207
9.08.3.1	Formation and Initial State	207
9.08.3.2	Thermal Evolution	207
9.08.3.2.1	Core cooling	207
9.08.3.2.2	Maintaining the geodynamo	208
9.08.3.2.3	Present-day energy budget	209
9.08.3.2.4	Thermal evolution	210
9.08.3.2.5	Inner-core age and initial core temperature	211
9.08.3.2.6	Consequences for the mantle	213
9.08.3.2.7	Stable stratification	213
9.08.3.3	Compositional Evolution	214
9.08.3.3.1	Inner-core growth	215
9.08.3.3.2	Core–mantle boundary	215
9.08.3.4	Summary	215
9.08.4	**Conclusions**	216
Acknowledgments		216
References		216

9.08.1 Introduction

The evolution of the Earth's core is important for three main reasons. First, the formation of the core was one of the central events in the ancient, but geologically rapid, period over which the Earth accreted and generated observational constraints on this poorly understood epoch. Second, the initial conditions, both thermal and compositional, established during this period have largely controlled the subsequent evolution of the core and may have also significantly affected the mantle. Finally, the evolution of the Earth's core resulted in the generation of a long-lived global magnetic field, which did not occur for the superficially similar cases of Mars or Venus. The objective of this chapter is to describe our current understanding of the evolution of the core, from shortly after its formation to the present day.

The first section of this chapter will summarize the present-day state of the core, since it is this state that is the end product of the core's evolution. The bulk of the chapter will then examine how the core evolved from its initial thermal and compositional state. Most of the arguments will be based on physics rather than chemistry, as compositional constraints on the core's long-term evolution are rare.

The material covered in this chapter follows on directly from the **Chapter 9.03**, in which the origin and formation of the core are discussed. Much of the discussion of the core's energy and entropy budgets is derived from a more thorough treatment in **Chapter 8.02**. Other aspects of the core's behavior are described in chapters in this Treatise by **Chapters 8.05**, **8.03**, and **8.11**. The companion *Treatise on Geochemistry* contains useful articles on planetary accretion (Chambers, 2003) and various aspects of core composition (Li and Fei, 2003; McDonough, 2003; Righter and Drake, 2003).

This chapter differs from the original *Treatise on Geophysics* chapter (Nimmo, 2007) in one main respect. Recent (2011–12) estimates of core thermal conductivity exceed earlier estimates by a factor ≈ 3 (**Section 9.08.2.2**). As a result, the heat fluxes required to drive a dynamo are larger, the inner core is likely a young feature, and regions of the core may have been stably stratified. This chapter reviews our current understanding in the light of these new results; readers are warned that certain topics

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00160-3

(especially the role of stratification) are not yet thoroughly understood and are likely to be the subject of intensive research in the next few years.

9.08.2 Present-Day State of the Core

Prior to investigating the earliest history and evolution of the core, it is important to briefly describe its present-day features. More detail can be found in the chapters referred to in the preceding text; here, I will focus on those parameters that are most important when considering the thermal and compositional evolution of the core. In particular, the uncertainties associated with these parameters will be assessed; doing so is important when assessing the likely range of thermal evolution outcomes (Section 9.08.3.2). The values and uncertainties adopted are discussed in the succeeding text and in more detail in Chapter 8.02. An important caveat to be borne in mind is that I generally assume a fully mixed (i.e., not stably stratified) core. A partially stratified core will experience different dynamo and thermal histories; the extent to which such an eventuality is likely is briefly discussed in Section 9.08.3.2.7.

9.08.2.1 Density and Pressure

The radially averaged density structure of the core may be derived directly from seismological observations (see Chapter 1.23 for reviews). The density of the core increases monotonically with depth, due to the increasing pressure. However, there is also a sharp density discontinuity at the inner-core boundary (ICB), which arises because of two effects. Firstly, solid core material is inherently denser than liquid core material at the same pressure and temperature (P, T) conditions. Secondly, the outer core contains more of one or more light elements than the inner core (e.g., McDonough, 2003; Poirier, 1994) and would therefore be less dense even if there were no phase change. This compositional density contrast $\Delta\rho_c$ has a dominant role in driving compositional convection in the core; unfortunately, its magnitude is uncertain by a factor ≈ 2.

The total density contrast across the ICB is somewhat uncertain. Seismological techniques suggest a range of about 5–8.3% (Cao and Romanowicz, 2004; Masters and Gubbins, 2003). The density contrast between pure solid and liquid Fe at the ICB is estimated at 1.6–1.8% (Alfe et al., 1999; Laio et al., 2000). These results imply a compositional density contrast of 3.2–6.7%, or $\triangle\rho_c \approx 400$–$800\ \mathrm{kg\ m^{-3}}$, and may in turn be used to estimate the difference in light element(s) concentrations between inner core and outer core, which helps sustain the dynamo (see Section 9.08.3.2.2).

For the theoretical models described later (Section 9.08.3), it is important to have a simple description of the density variation within the Earth. One such description is given by Labrosse et al. (2001), where the variation of density ρ with radial distance r from the center of the Earth is given by

$$\rho(r) = \rho_{cen}\exp\left(-r^2/L^2\right) \quad [1]$$

where ρ_{cen} is the density at the center of the Earth and L is a length scale given by

$$L = \sqrt{\frac{3K_0(\ln(\rho_{cen}/\rho_0)+1)}{2\pi G\rho_0\rho_{cen}}} \quad [2]$$

Here, K_0 and ρ_0 are the compressibility and density at zero pressure, respectively; G is the universal gravitational constant; and $L = 7272$ km using the parameters given in Chapter 8.02. Although this expression neglects the density jump at the ICB, the error introduced is negligible compared to other uncertainties.

The corresponding pressure is given by

$$P(r) = P_c + \frac{4\pi G\rho_{cen}^2}{3}\left[\left(\frac{3r^2}{10} - \frac{L^2}{5}\right)\exp\left(-r^2/L^2\right)\right]_r^R \quad [3]$$

where P_c is the pressure at the core–mantle boundary (CMB) and R is the radius of the core.

9.08.2.2 Thermodynamic Properties

From the point of view of the thermal evolution of the core, the most important parameters are those that determine the temperature structure and heat flux within the core, in particular the thermal conductivity k and expansivity α (see Section 9.08.2.4).

As discussed in Chapter 8.02, recent determinations of core thermal conductivity differ significantly from the traditional estimate of about $46\ \mathrm{W\ m^{-1}\ K^{-1}}$ (Stacey and Anderson, 2001). Two molecular dynamics calculations (De Koker et al., 2012; Pozzo et al., 2012) have been confirmed by high-pressure experimental measurements (Gomi et al., 2013) and yield thermal conductivities of 100–$150\ \mathrm{W\ m^{-1}\ K^{-1}}$ at the CMB, depending on the assumed temperature and composition, and up to $250\ \mathrm{W\ m^{-1}\ K^{-1}}$ at the center of the core. The depth dependence of the conductivity plays an important role in determining which regions of the core are stably stratified. For simplicity, I will generally assume a constant value of k (except when considering stable stratification) (Section 9.08.3.2.7) and take the baseline value to be $130\ \mathrm{W\ m^{-1}\ K^{-1}}$.

The thermal expansivity within the core may be obtained from seismology if the Grüneisen parameter is known. This parameter likely remains constant at roughly 1.5 throughout the core (e.g., Alfè et al., 2002b; Anderson, 1998). Because the seismic parameter increases with depth, α increases by a factor of 1.5–2 from the center of the Earth to the CMB (Labrosse, 2003; Roberts et al., 2003). However, little accuracy is sacrificed if a constant mean value of α is adopted, furthermore, assuming that both k and α are constant results in adiabatic temperature profiles that closely resemble those in which both parameters vary. I therefore adopt a range for α of 0.8–$1.9 \times 10^{-5}\ \mathrm{K^{-1}}$. A list of estimated values for important parameters is given in Table 2; more details may be found in Chapter 8.02.

9.08.2.3 Composition

The composition of the core is important because it potentially provides constraints on its origin and mode of formation. Unfortunately, as will be seen in the succeeding text, the constraints provided are currently rather weak, as few elements have well-known core abundances.

9.08.2.3.1 Light elements

It is clear from seismology and experiments that the outer core is 6–10% less dense than pure liquid iron would be under the estimated P, T conditions (e.g., Alfè et al., 2002a). While the core almost certainly contains a few weight percent of nickel (e.g., McDonough, 2003), this metal has an almost identical density to iron and is thus not the source of the density deficit (e.g., Li and Fei, 2003). The inner core also appears to be less dense than a pure iron composition would suggest (Jephcoat and Olson, 1987), though here, the difference is smaller. Both the outer core and inner core must therefore contain some fraction of light elements, of which the most common suspects are sulfur, silicon, oxygen, carbon, and hydrogen (see Hillgren et al., 2000; Li and Fei, 2003; Poirier, 1994 for reviews). For any particular element, or mixture of them, the inferred density deficit may be used to infer the molar fraction of the light element(s) present. Because the density deficit is larger in the outer core, it is thought that light elements are being expelled during crystallization of the inner core. This expulsion is of great importance, because it generates compositional convection that helps drive the geodynamo (see **Section 9.08.3.2.2**).

Apart from their role in driving the dynamo, these light elements are important for two other reasons. Firstly, they probably reduce the melting temperature of the core by several hundred degrees Kelvin (see **Section 9.08.2.4**). Secondly, if the actual light elements in the core could be reliably identified, they would provide a strong constraint on the conditions (e.g., oxidation state) under which the core formed. As will become clear, there is not yet a consensus on which elements are in fact present. Fortunately, from a geophysical point of view, this does not matter; what matters is the density contrast across the ICB, which is known from seismology.

Table 1 gives several examples of model core compositions. All these models are derived by comparing estimates of the bulk silicate Earth elemental abundances (inferred from upper-mantle nodules and crustal samples), with estimates of the initial solar nebular composition (based mainly on chondritic meteorite samples). Although there is some agreement on the abundances of Fe, Ni, and Co, the relative abundances of the light elements (Si, S, O, and C) vary widely. The abundance of H in the core cannot be modeled in this way because of its extreme volatility; in practice, it will be determined by the P, T conditions and amount of H in the Earth's mantle prior to and during differentiation (see, e.g., Abe et al., 2000; Okuchi, 1997).

These cosmochemical models do not take into account the ease with which different elements partition into iron under the relevant conditions. Unfortunately, because partitioning is a strong function of pressure, temperature, and oxygen fugacity, models of core composition taking partitioning into account also have to make assumptions about the conditions under which core formation happened. As a result, some calculations favor Si as the dominant element (Ricolleau et al., 2011; Rubie et al., 2011), while others prefer O (Siebert et al., 2013). Low concentrations (<2.5 wt%) of O in the outer core are also implied by comparison of experimental velocity and density determinations with seismologically derived values (Huang et al., 2011).

On the other hand, the presence of O in the outer core is favored, based in its inferred ability to be expelled during the inner-core growth. The outer core appears to contain more of the light element(s) than the inner core. This implies that one or more of the light elements must partition strongly into liquid iron during freezing, which is potentially a diagnostic behavior. Oxygen, due to its small atomic radius, tends to be expelled during freezing (Alfè et al., 2002a). Conversely, S and Si have atomic radii similar to that of iron at core pressures and thus substitute freely for iron in the solid inner core. These results thus support the case for O being one of the light elements. Further work is needed to resolve these apparently contradictory results.

Although not traditionally viewed as a light element, Stevenson (2012) had pointed out that the apparently strong temperature dependence of Mg solubility in iron (Nomura et al., 2012) may have important consequences for dynamo evolution. If metal–silicate equilibration happened at high-temperature conditions, as, for instance, during the Moon-forming impact, then the core will have initially contained high concentrations of MgO. As the core cooled, this MgO will have exsolved throughout the core and risen to form a buoyant layer at the top of the core. A similar proposal for Si was previously made by Gessmann et al. (2001). Such a stably stratified layer has previously been proposed on other grounds (see **Section 9.08.3.3.1**); the importance of this proposal is that exsolution provides a potent mechanism to drive the dynamo, because of the large density contrast involved. Such a scenario could potentially allow a long-lived dynamo without requiring excessively high core heat fluxes.

Table 1 Model core compositions

	MA	*WD*	*A+*	*McD-1*	*McD-2*
Fe wt%	84.5	80.3	79.4	85.5	88.3
Ni	5.6	5.5	4.9	5.2	5.4
Si	–	14.0	7.4	6.0	–
S	9.0	–	2.3	1.9	1.9
O	–	–	4.1	–	3.0
C	–	–	–	0.2	0.2
Co	0.26	0.27	0.25	–	–

MA, Morgan and Anders (1980); WD, Wanke and Dreibus (1988); A+, Allegre et al. (1995); McD-1 and McD-2 refer to two different models given in McDonough (2003).

9.08.2.3.2 Isotopes

Although the bulk of the core consists of Fe, Ni, and a few weight percent of light elements, some trace elements are also of importance. Firstly, the radioactive isotopes of K, Th, and U can potentially have a significant effect on both the age of the inner core and the maintenance of the dynamo (e.g., Buffett, 2002; Labrosse et al., 2001; Nimmo et al., 2004). Unfortunately, there is as yet little agreement on whether or not such elements are really present in the core. Longer discussions on this issue may be found in McDonough (2003) and Roberts et al. (2003); only a brief summary is given here, while the role of radioactive elements in core evolution is discussed in **Section 9.08.3.2**.

There is little evidence, from either cosmochemistry or partitioning experiments, to expect either U or Th to partition into the core. On the other hand, the Earth's mantle is clearly depleted in K relative to chondrites (e.g., Lassiter, 2004). However, since K is a volatile element, it is unclear whether this depletion is due to sequestration of K in the core or simple loss

of K from the Earth as a whole early in its history. Experimental investigations (Bouhifd et al., 2007; Corgne et al., 2007; Gessmann and Wood, 2002; Lee et al., 2004; Murthy et al., 2003) showed that partitioning of K into core materials is possible but depends in a complex fashion on other factors such as the amount of sulfur or oxygen present. The highest-pressure experiments to date (Nomura et al., 2012) suggest minimal quantities of potassium in the core, but at these conditions, the compositional parameter space has yet not been thoroughly explored. The removal of K from the core would also likely involve the removal of other elements with similar affinities for iron, but it is not yet clear what constraints the observed abundances of these other elements place on the amount of K in the core. It currently appears that up to a few hundred ppm K in the core are permitted, but not required, by both the experiments and the geochemical observations. The detection of antineutrinos produced by radioactive decay in the Earth's interior (Gando et al., 2011) may help ultimately resolve this question.

Certain isotopic systems constrain the timing and P, T conditions of core formation. The Hf–W (Kleine et al., 2009), U–Pb (Wood and Halliday, 2005), and Pd–Ag (Schonbachler et al., 2010) systems provide constraints on core formation timescales and are discussed in some detail in **Chapter 9.03**. Stable isotopes of Fe and Si may exhibit small degrees of fractionation as a result of core formation (e.g., Georg et al., 2007; Polyakov, 2009), though the evidence is currently controversial.

A final set of potentially important isotopes are those of rhenium, because they may constrain the onset of inner-core formation (e.g., Walker et al., 1995). As summarized in **Section 9.08.3.3.1**, this hypothesis has been largely rejected; nonetheless, it is mentioned here for completeness.

9.08.2.4 Temperature Structure

Both the temperature structure within the core and the shape of the melting curve play an important role in determining the thermal evolution of the core. Assuming that the entire liquid core is convecting (i.e., no stable stratification), its mean temperature profile will be that of an adiabat, except at the very thin top and bottom boundary layers. Since the temperature at the ICB must equal the melting temperature of the core at that pressure (**Figure 1**), the temperature elsewhere in the core may be extrapolated from the ICB conditions by using the appropriate adiabat. Thus, determining the melting behavior of core material is crucial to establishing the temperature structure of the core.

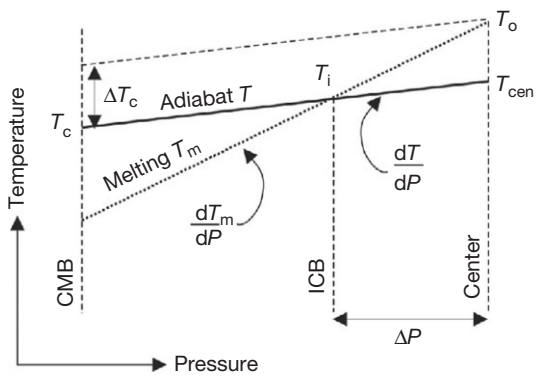


Figure 1 Definition sketch for temperature structure within the core. Note that in reality, neither temperature profile is linear.

The adiabatic temperature T within the core is given by (Labrosse et al., 2001)

$$T(r) = T_{cen}\exp\left(-r^2/D^2\right) \quad [4]$$

where T_{cen} is the temperature at the center of the Earth and D is a length scale given by

$$D = \sqrt{3C_p/2\pi\alpha\rho_{cen}G} \quad [5]$$

Here, C_p is the specific heat capacity, α is the thermal expansivity (**Section 9.08.2.2**), and $D = 6203$ km using the parameters given in **Table 2**, model 2.

As discussed in more detail in **Chapter 8.02**, first-principles calculations have been used to determine the melting curve of iron and its dependence on impurities. Alfè et al. (2003) predicted a temperature at the ICB T_i of 5650 ± 600 K, taking into account the reduction in temperature due to the light element(s). The gradient in melting temperature is roughly 9 K GPa^{-1} at the ICB. These results are broadly consistent with the low-pressure diamond anvil cell results of Shen et al. (1998) and Ma et al. (2004), though not those of Boehler (1993). Similarly, the results agree with the higher-pressure shock-wave results of Brown and McQueen (1986) and Nguyen and Holmes (2004), though not those of Yoo et al. (1993). The numerical results of Belonoshko et al. (2000) and Laio et al. (2000) also give similar answers once corrections due to the different molecular dynamics techniques used have been applied. Further discussions of the differing results and the reliability of different theoretical approaches may be found in Alfè et al. (2004) and Bukowinski and Akber-Knutson (2005).

Given the ICB temperature and the relevant thermodynamic quantities, the temperature at the core side of the CMB, T_c, may be deduced and is here taken to be 4180 K. This value is actually of only secondary importance as far as the thermal evolution of the core is concerned; of much greater interest are the relative slopes of the adiabat and the melting curve (see the succeeding text).

Although other approaches may be adopted (e.g., Buffett et al., 1996; Labrosse et al., 2001; Roberts et al., 2003), a reasonable approximation is that the core melting temperature is a quadratic function of pressure. That is, the core melting temperature T_m may be written as

$$T_m(P) = T_{m0}\left(1 + T_{m1}P + T_{m2}P^2\right) \quad [6]$$

where T_{m0}, T_{m1}, and T_{m2} are constants and T_{m0} incorporates the reduction in melting temperature due to the light element(s). Note that this parameterization is only intended to apply over the pressure range relevant to the core, so that T_{m0} should not be taken to be the zero-pressure melting temperature.

When considering the growth history of the inner core, the crucial parameter is the difference in gradients between the melting curve and the adiabat. One way of expressing this quantity is to define ΔT_c, the change in the CMB temperature since the onset of inner-core solidification. As shown

Table 2 Parameter values adopted (eq. is the equation or section in which each parameter is defined)

	Model						Model				
	1	*2*		*Units*	*eq.*		*1*	*2*		*Units*	*eq.*
k	90	130		W m^{-1} K^{-1}	Section 9.08.2.2	$\Delta\rho_c$	800	560		kg m^{-3}	Section 9.08.2.1
ΔT_c	150	60		K	[7]	α	0.9	1.25	$\times10^{-5}$	K^{-1}	[5]
D	7310	6203		km	[5]	T_{cen}	5243	5726		K	[4]
T_i	5099	5509		K	Section 9.08.2.4	T_o	5431	5808		K	Section 9.08.2.4
T_{m0}	8034	2677		K	[6]	T_{m1}	−3.38	2.95	$\times10^{-12}$	Pa^{-1}	[6]
T_{m2}	8.37	69	$\times10^{-25}$	Pa^{-2}	[6]	$\tilde{E}_T$	3.2	2.0	$\times10^{23}$	J/K^2	[10]
Q_k	7.5	15.0		TW	[12]	$\tilde{Q}_T$	4.6	3.3	$\times10^{27}$	J/K	[8]

Model 1 is an end-member case using parameter values designed to generate an ancient inner core; model 2 is the best guess at the real parameter values. Other parameters not specified here are assumed constant in both models and are generally the same as those adopted in **Chapter 8.02**. In particular, $T_c = 4180$ K, latent heat $L_H = 750$ kJ kg^{-1}, $C_p = 840$ J kg^{-1} K^{-1}, $\rho_{cen} = 12\,500$ kg m^{-3}, and the melting gradient at the ICB is 9.4 K GPa^{-1}.

in Figure 1, when the two curves are linear, ΔT_c may be defined as follows

$$\Delta T_c = \frac{\Delta P}{f_{ad}}\left(\frac{dT_m}{dP} - \frac{dT}{dP}\right) = 23K\left(\frac{(dT_m/dP) - (dT_a/dP)}{1\,\mathrm{K/GPa}}\right)\left(\frac{1.32}{f_{ad}}\right) \quad [7]$$

where ΔP is the pressure difference between the present ICB and the center of the Earth; T_a and T_m are the adiabatic and melting temperatures, respectively; f_{ad} is a factor converting the temperature at the CMB to that at the ICB; and the numerical value comes from the parameters of model 2 (Table 2). By choosing values for T_c, ΔT_c, and the adiabatic gradient, the ICB temperature T_i and dT_m/dP are then determined (see Table 2). In the case that the melting curve exhibits curvature (eqn [6]), one also has to specify dT_m/dP at the ICB in order to solve for T_{m0}, T_{m1}, and T_{m2}.

Because the adiabatic and melting gradients are both uncertain and of similar sizes, the uncertainty in ΔT_c tends to be amplified. Values for ΔT_c from four studies (Buffett et al., 1996; Labrosse, 2003; Pozzo et al., 2012; Roberts et al., 2003) range from 43 to 146 K, with smaller values implying younger inner cores (see **Chapter 8.02**). Here, I will assume a range of 50–150 K as representative of the likely uncertainties. Taking the melting gradient at the ICB to be 9.4 K/GPa, Table 2 lists the resulting quantities for a best-guess set of parameters (model 2) and a set of parameters designed to maximize inner-core age (model 1).

9.08.2.5 Inner-Core Structure

As reviewed in **Chapters 1.23** and **8.12**, the inner core displays a surprising amount of structure. It has a depth-dependent anisotropy and also exhibits differences in seismic properties between the eastern and western hemispheres. The causes of these observations are currently uncertain: possible explanations include magnetic stresses, inner-core convection, and growth/translation mediated by spatial variations in heat loss at the ICB. This inner-core structure may ultimately provide a constraint on how the inner core has grown, but our present understanding is too limited to be useful.

9.08.2.6 The CMB Region

The CMB region is relevant to core evolution for two reasons. Firstly, it is the behavior of this region, and in particular its rheology and temperature structure, which controls the rate at which heat is extracted from the core. As a consequence, the thermal evolution of the core is intimately tied to that of the mantle. This interdependence between the core and mantle is one of the reasons that theoretical investigations of core thermal evolution are subject to such large uncertainties.

Secondly, the CMB region is interesting from a compositional point of view, since it is the point at which regions containing elements with very different chemical potentials meet. The extent to which the resulting reactions have influenced the behavior of the deepest mantle (or the outermost core) is unclear. However, because these reactions provide indications of core evolution, the question of whether core or CMB material can plausibly be entrained to the surface is of considerable interest (**Section 9.08.3.3.2**).

Whether or not a dynamo can be sustained ultimately depends on the CMB heat flow, that is, the rate at which heat is extracted from the core (**Section 9.08.3.2.2**). The CMB heat flow, in turn, is determined by the ability of the mantle to remove heat. Importantly, independent estimates on this cooling rate exist, based on our understanding of mantle behavior.

As reviewed by Tackley (2012) and Lay et al. (2008), the CMB region is complicated, because lateral variations in bulk chemistry, phase, and melt fraction apparently all occur. Nonetheless, several different techniques have been used to estimate the heat flux across the CMB, with broadly similar answers.

The simplest approach is to assume conductive heat flow across the seismologically determined bottom boundary layer (termed D″). The temperature above this boundary layer can be determined by downward extrapolation of the mantle adiabat, while the temperature below is just T_c. As discussed in **Chapter 8.02**, for likely lower-mantle thermal conductivities, the resulting conductive heat flow is probably in the range 8–16 TW. This simple estimate, however, may neglect important complications, such as heat production within D″. Wu et al. (2011) used a more sophisticated statistical approach to infer a range of 13 ± 3 TW. An alternative is to compare

convection models with observations of mantle plumes to infer the CMB heat flow. For instance, Leng and Zhong (2008) found a value of ≈11 TW.

Seismological detection of the deep-mantle phase transition from perovskite to postperovskite has been used to obtain a more direct measurement of temperature gradient and thus heat flow. Two suitable regions yield extrapolated global CMB heat flows of 9–17 TW (Lay et al., 2006) and 7–15 TW (Van der Hilst et al., 2007). Nakagawa and Tackley (2010) pointed out that these regions have higher than average heat flux, so that the globally averaged flow may be lower (around 9 TW).

One additional complication is that D″ may contain enhanced concentrations of radioactive elements (e.g., Boyet and Carlson, 2005; Coltice and Ricard, 1999; Tolstikhin and Hofmann, 2005). In this case, the heat flow out of the core will be less than the heat flows inferred from the estimates quoted in the preceding text (Buffett, 2002; Labrosse et al., 2007). Bearing this caveat in mind, a range of 12 ± 5 TW is likely to encompass the real present-day CMB heat flow. This range of heat flows suggests a current core cooling rate dT_c/dt of 50–115 K Gy^{-1}, using the parameters for model 2.

The compositional nature of the CMB region may also have an effect on the evolution of the core. The CMB region may be at least partially molten, an inference supported by the presence of a (laterally discontinuous) ultralow-velocity zone (e.g., Garnero et al., 1998). The presence of such a melt layer, which is probably denser than the surrounding solid material, is likely to affect heat transfer from the core to the mantle (Labrosse et al., 2007). Such a layer is also likely to have been more extensive in the past, when core temperatures were higher (see **Section 9.08.3.2.5** and **Chapter 9.03**). Another possibility is the presence of high-density, compositionally distinct material, probably subducted oceanic crust. Again, this material, especially if enriched in radioactive materials (Buffett, 2002), is likely to have affected long-term core evolution (Nakagawa and Tackley, 2004). Finally, the effect of CMB dynamics on the perovskite phase transition is not yet clear; for instance, if the viscosity of postperovskite is low, the heat flow across the CMB will be increased (Nakagawa and Tackley, 2011). The manner in which the CMB may have evolved with time in response to the evolution of the core is discussed further in **Section 9.08.3.3.2**.

9.08.2.7 Dynamo Behavior Over Time

One might expect that the behavior of the Earth's magnetic field over time would provide information on the evolution of the dynamo and core. However, despite much work on this subject (see reviews by Jacobs (1998) and Valet (2003) and **Chapter 5.13**), the information is limited to the following: (1) A reversing, predominantly dipolar field has existed, at least intermittently, for at least the last 3.5 Gy; (2) the intensity of the field does not appear to have changed in a systematic fashion over time.

There are several reasons why there are so few constraints. First, the magnetic field that we can measure at the surface is different in both frequency content and amplitude from the field within the core. In particular, ohmic heating is dominated by small-scale magnetic fields, which are not observable at the surface (see the succeeding text). Second, the number of observations on paleomagnetic fields declines dramatically prior to ≈150 My BP because of the almost complete absence of unsubducted oceanic crust. Third, there is little theoretical understanding of how changes in core behavior relate to changes in the observed magnetic field.

The first two problems are unlikely to be resolved in the foreseeable future. However, there has been some progress with the third, thanks to increasingly realistic simulations of the geodynamo (see **Chapter 8.10**). For instance, a study by Roberts and Glatzmaier (2001) found that increasing the inner-core size tended to result in a less axisymmetric field and (surprisingly) greater time variability. Thus, at least in theory, observed changes in the time variability of the magnetic field with time could be used to place constraints on the evolution of the Earth's core. Similarly, the degree to which the field departs from dipolarity may provide a useful metric (e.g., Pavlov and Gallet, 2010; Smirnov et al., 2011). Any apparent variation in the field amplitude with time (e.g., Labrosse and Macouin, 2003), however, is less likely to be useful: Bloxham (2000), Roberts and Glatzmaier (2001), and Aubert et al. (2009) all found that the magnetic field as observed at the surface of the Earth was relatively insensitive to the existence or size of an inner core. Furthermore, it seems likely that mantle behavior may have at least as large an effect on the geodynamo as changes in core variables such as cooling rate or inner-core size (Biggin et al., 2012; Glatzmaier et al., 1999; Olson et al., 2010).

In spite of the difficulties in extracting detailed information on core evolution from the paleomagnetic record, an important result is that the geodynamo appears to have persisted, without long-term interruptions, for at least 3.5 Gy (McElhinny and Senanayake, 1980; Tarduno et al., 2010). The earliest documented apparent paleomagnetic reversal is at 3.2 Gy BP (Layer et al., 1996). Although the amplitude of the field has varied with time (Prevot et al., 1990; Selkin and Tauxe, 2000), the maximum field intensity appears never to have exceeded the present-day value by more than a factor of five (Dunlop and Yu, 2004; Valet, 2003), while Archean magnetic field strengths were comparable to present-day values (Tarduno et al., 2010).

In summary, the fact that a reversing dynamo has apparently persisted for >3.5 Gy can be used to constrain the evolution of the core over time (see **Section 9.08.3.2**). Unfortunately, other observations, which might potentially provide additional constraints, such as the evolution of the field intensity, are either poorly sampled or difficult to relate to the global energy budget or both.

9.08.2.7.1 Ohmic dissipation

As discussed in the succeeding text, the power dissipated in the core by ohmic heating is a critical parameter to determining whether a dynamo can operate: a more dissipative dynamo requires more rapid core cooling and a higher CMB heat flux. Unfortunately, this heating rate is currently very poorly constrained. The heating is likely to occur at length scales that are sufficiently small that they can be neither observed at the surface nor resolved in numerical models (Roberts et al., 2003). Moreover the toroidal field, which is undetectable at the surface, may dominate the heating.

The ohmic dissipation Q_Φ may be converted to an entropy production rate E_Φ using $E_\Phi = Q_\Phi/T_D$ (Roberts et al., 2003), where the characteristic temperature T_D is unknown but intermediate between T_i and T_c and is here assumed to be 5000 K. The entropy production rate is simply a convenient way of assessing the potential for generating a dynamo and is discussed in more detail in Section 9.08.3.2.2 and Chapter 8.02. One approach to estimating the required rate is to extrapolate from numerical dynamo simulations. Roberts et al. (2003) used the results of the Glatzmaier and Roberts (1996) simulation to infer that 1–2 TW is required to power the dynamo, equivalent to an entropy production rate of 200–400 MW K^{-1}. Buffett (2002) suggested 0.1–0.5 TW and Buffett and Christensen (2007) 0.7–1 TW. Christensen and Tilgner (2004) gave a range of 0.2–0.5 TW, but this range was later updated to 1.1–15 TW (Christensen, 2010) or 3–8 TW (Stelzer and Jackson, 2013). There is clearly little consensus on the dissipation rate required, but a few TW (equivalent to a few hundred MW K^{-1}) seem reasonable. Labrosse (2003) argued for a range 350–700 MW K^{-1}, and Gubbins et al. (2003) favored 500–800 MW K^{-1}. We shall regard the required ohmic dissipation rate as currently unknown but think it likely that entropy production rates in excess of 100 MW K^{-1} are sufficient to guarantee a geodynamo.

9.08.2.8 Summary

The present-day temperature structure and composition of the core establish boundary conditions, which constrain both the core's initial mode of formation and subsequent evolution. In particular, the size of the inner core and the persistence of the geodynamo for at least 3.5 Gy place constraints on the CMB heat flux. Light elements in the Earth's core not only help to power the dynamo but also constrain the conditions under which the Earth formed.

The next section will examine how the core evolved from the initial conditions established by the accretion process to its inferred present-day state.

9.08.3 Evolution of the Core

9.08.3.1 Formation and Initial State

The initial thermal and chemical conditions of the Earth's core were determined by the manner in which the Earth accreted, a relatively geologically rapid process. This period of the core's history is discussed extensively in Chapter 9.03, and only a brief summary is given here.

Theoretical arguments and geochemical observations suggest that the Earth accumulated the bulk of its mass through a few, large impacts within about 100 My of solar system formation and that each of these impacts generated a near-global, possibly transient, magma ocean. Although the impacting bodies were undoubtedly differentiated, preexisting chemical signals appear to have been overprinted by the impact process. The impactor cores likely underwent some degree of emulsification as they traversed the magma ocean, resulting in partial chemical reequilibration. The potential energy liberated as the metal sank resulted in high initial core temperatures and likely extensive melting in the lowermost mantle.

After this initial period of large and geologically rapid transfers of mass and energy, the subsequent thermal and compositional evolution of the core – the focus of this section – was much less dramatic. Unfortunately, there are few observational constraints on the details of this longer-term evolution. As discussed in the succeeding text, present-day observations (in particular, the size of the inner core and estimates of the CMB heat flux) provide some constraints. The fact that a geodynamo has apparently operated for at least 3.5 Gy provides a lower bound on the rate at which the core must have cooled. However, it is important to note that the long-lived field does not necessarily require a similarly ancient inner core.

The first half of this section will investigate the thermal evolution of the core. In particular, it will focus on three questions: How much has the core cooled over time, when did the inner core start to grow, and how was the dynamo maintained? Because of the paucity of observational constraints, this section will focus on theoretical approaches and in particular on the uncertainties introduced by uncertainties in the relevant parameters. The second half will focus on the compositional evolution of the core, in particular the chemical effects of inner-core formation, and possible reactions taking place at the CMB.

9.08.3.2 Thermal Evolution

The thermal evolution of the Earth's core has been the subject of considerable interest over the last decade. As outlined in Section 9.08.2, experimental uncertainties have led different groups to adopt different values for parameters of interest, such as the thermal conductivity. Accordingly, the calculations carried out in this section will make use of two different sets of parameters (Table 2): one end-member designed to maximize the likelihood of an ancient inner core (model 1) and one using the best-guess parameter values (model 2). In this way, the uncertainties involved in the theoretical calculations will be made clear, while conclusions that are robust under both models are likely to prove durable.

9.08.3.2.1 Core cooling

In one sense, the thermal evolution of the core is relatively simple. Heat is extracted out of the core at the CMB, at a rate that depends primarily on processes within the mantle. As a result, in the absence of an internal heat source, the core cools with time. At some point, the core adiabat crosses the melting curve, and inner-core solidification begins (Figure 1).

The instantaneous energy balance within the core may be written (e.g., Buffett et al., 1996; Gubbins et al., 2003; Roberts et al., 2003; Chapter 8.02)

$$Q_{cmb} = Q_s + Q_g + Q_L + Q_R = \widetilde{Q}_T \frac{dT_c}{dt} + Q_R \qquad [8]$$

Here, Q_{cmb} is the heat flow across the CMB; the core contributions Q_s, Q_g, Q_L, and Q_R are, respectively, from secular cooling, gravitational energy release, latent heat release, and radioactive decay; and the outer core is assumed to be adiabatic and homogeneous. Note that the assumption that the outer core is well mixed and convecting throughout may not be the case if a stable conductive layer (e.g., Gomi et al., 2013; Labrosse et al., 1997; Pozzo et al., 2012) or a compositionally

buoyant layer (e.g., Braginsky, 2006) is present at either the top or the bottom of the core. In what follows, I will generally assume that the core is well mixed and adiabatic; the potential role of a stably stratified region will be addressed in Section 9.08.3.2.7. I will also neglect the potential role of exsolution (Section 9.08.2.3.1) since it is as yet poorly quantified.

The first three terms in eqn [8] are all proportional to the core cooling rate dT_c/dt, where T_c is the core temperature at the CMB and $\widetilde{Q}_T$ is a measure of the total energy released per unit change in core temperature. Both Q_g and Q_L depend on the inner-core size and are zero in the absence of an inner core. This equation allows the evolution of the core temperature to be calculated if the CMB heat flux through time is known.

This energy balance has several important consequences. Firstly, when inner-core formation begins, the same CMB heat flux results in a reduced core cooling rate, because of the extra energy terms (Q_g, Q_L). Secondly, the result of radioactive heating is likewise to reduce the core cooling rate for the same CMB heat flux.

Radiogenic elements can have a strong effect on the core cooling rate and thus the age of the inner core (e.g., Labrosse et al., 2001). Rewriting eqn [8], we obtain a core cooling rate of

$$\frac{dT_c}{dt} = \frac{Q_{cmb} - Q_R}{\widetilde{Q}_T} \qquad [9]$$

It is clear that the effect of the Q_R term is to reduce the rate of core cooling and hence prolong the life of the inner core. This is an issue we return to in the succeeding text.

A major disadvantage with eqn [8] is that it does not include an ohmic dissipation term, because the transformation of kinetic energy to magnetic energy to heat occurs without changing the global energy balance (see Gubbins et al., 2003). This equation is therefore not useful in determining the evolution of the geodynamo.

9.08.3.2.2 Maintaining the geodynamo

How the geodynamo is maintained has been a question of considerable interest since the initial work of Bullard (1950), Verhoogen (1961), and Braginsky (1963). The entropy balance approach described in the succeeding text was developed in the 1970s (Backus, 1975; Gubbins, 1977; Gubbins et al., 1979; Hage and Muller, 1979; Hewitt et al., 1975; Loper, 1978) and has been the focus of ongoing activity (Braginsky and Roberts, 1995; Buffett, 2002; Buffett et al., 1996; Gubbins et al., 2003, 2004; Labrosse, 2003; Lister, 2003; Lister and Buffett, 1995; Pozzo et al., 2012; Roberts et al., 2003).

Ultimately, the geodynamo is maintained by the work done on the field by convective motions. This convection is driven partly by the extraction of heat into the overlying mantle and partly by the fact that the resulting inner-core growth (or exsolution) releases light elements into the base of the outer core. Thus, both thermal and compositional convection are important, with the relative contributions depending on the different parameter values adopted, in particular the size of the inner core.

Just as eqn [8] describes the energy balance in the core, an equivalent equation can be derived for the entropy balance (e.g., Gubbins et al., 2003, 2004; Labrosse, 2003; Lister, 2003; Roberts et al., 2003; Chapter 8.02). The latter equation does include ohmic dissipation (dissipation is nonreversible and is thus a source of entropy). The entropy may be thought of as the power divided by a characteristic temperature and multiplied by a thermodynamic efficiency factor. Different mechanisms (e.g., thermal and compositional convection) have different efficiency factors (e.g., Buffett et al., 1996; Lister, 2003). Unfortunately, it is not currently understood how to relate the entropy production rate to global magnetic field characteristics, such as reversal frequency (Section 9.08.2.7).

The entropy rate available to drive the dynamo may be written as (e.g., Gubbins et al., 2004; Labrosse, 2003)

$$\Delta E = E_s + E_L + E_g + E_H + E_R - E_k = \widetilde{E}_T \frac{dT_c}{dt} + E_R - E_k \qquad [10]$$

where E_s, E_L, E_g, and E_H are the contributions due to cooling, latent heat release, gravitational energy release, and heat of reaction, respectively; E_R depends on the presence of radioactive elements in the core; and E_k depends on the adiabatic heat flux at the CMB. The first four terms are all proportional to the core cooling rate dT_c/dt, and $\widetilde{E}_T$ is simply a convenient way of lumping these terms together. This equation illustrates two important points. Firstly, as expected, a higher cooling rate or a higher rate of radioactive heat production increases the entropy rate available to drive a dynamo. Secondly, a larger adiabatic contribution (e.g., higher thermal conductivity) reduces the available entropy.

By combining eqns [8] and [10], an expression may be obtained that gives the core heat flow required to sustain a dynamo characterized by a particular entropy production rate E_Φ:

$$Q_{cmb} = Q_R \left(1 - \frac{T_T}{T_R}\right) + T_T(E_\Phi + E_k) \qquad [11]$$

where T_R is the effective temperature such that $T_R = Q_R/E_R$ and likewise $T_T = \widetilde{Q}_T/\widetilde{E}_T$. This equation encapsulates the basic physics of the dynamo problem.

Equation [11] shows that larger values of adiabatic heat flow or ohmic dissipation require a correspondingly higher CMB heat flow to drive the dynamo, as would be expected. In fact, in the absence of radiogenic heating, the CMB heat flow required is directly proportional to $E_k + E_\Phi$. The constant of proportionality depends on the thermodynamic efficiency of the core, which increases if an inner core is present. Because the term $(1 - (T_T/T_R))$ exceeds zero, a dynamo that is partially powered by radioactive decay will require a greater total CMB heat flow than the same dynamo powered without radioactivity. Alternatively, if the CMB heat flow stays constant, then an increase in the amount of radioactive heating reduces the entropy available to power the dynamo.

Equation [11] also illustrates the fact that a dissipative dynamo can exist even if the CMB heat flow is less than that conducted along the adiabat (i.e., subadiabatic) (Loper, 1978). In the absence of radioactivity, the entropy production rate E_Φ available for the dynamo is $(Q_{cmb}/T_T) - E_k$, which for the present-day core exceeds zero unless Q_{cmb} is strongly subadiabatic. Thus, a subadiabatic CMB heat flow can sustain a dynamo, as long as an inner core is present to drive compositional convection (e.g., Labrosse et al., 1997; Loper, 1978).

Although it is often assumed that compositional convection or lateral heat flux variations can keep the entire core well mixed despite the subadiabatic heat flux, this situation may lead to a stably stratified region within the core; such an eventuality is discussed further in Section 9.08.3.2.7.

In the absence of an inner-core and radiogenic heating, it may be shown that

$$Q_{cmb} = Q_k\left(1 + \frac{E_\Phi}{E_k}\right) \quad [12]$$

This equation shows that the heat flow at the CMB Q_{cmb} must exceed the adiabatic heat flow Q_k for a dynamo driven only by thermal convection to function. This result is important, because it demonstrates that there is no problem with sustaining a dynamo prior to the onset of inner-core formation, as long as the core cooling rate (or CMB heat flux) is large enough. This equation also allows dynamo dissipation to be taken into account explicitly: a more strongly dissipative core dynamo requires a more superadiabatic CMB heat flow to operate.

9.08.3.2.3 Present-day energy budget

Figure 2 shows how the rate of entropy production available to drive a dynamo varies as a function of the heat flow out of the core, both for a set of core parameters appropriate to the present-day Earth (with and without potassium) and for a situation in which the inner core has not yet formed. Figure 2(a) illustrates the case for the best-guess parameters (model 2), while Figure 2(b) uses parameters designed to maximize the inner-core age (model 1). As expected, higher core heat fluxes generate higher rates of entropy production; also, the same cooling rate generates more excess entropy when an inner core exists than when thermal convection alone occurs.

As discussed in the preceding text, when an inner core is present, positive contributions to entropy production arise from core cooling, latent heat release, and gravitational energy; the adiabatic contribution is negative (eqn [10]). For a present-day, potassium-free core, CMB heat flows of <6.5 and <2.5 TW result in negative entropy contributions and, therefore, no dynamo for models 2 and 1, respectively. Higher core cooling rates generate a higher net entropy production rate; they also mean that the inner core must have formed more recently.

For the present-day estimated CMB heat flow of 7–17 TW (Section 9.08.2.6), the net entropy production rate available to drive the dynamo is 50–850 MW K^{-1}, sufficient to generate roughly 0.3–4 TW of ohmic dissipation. Since most estimates of ohmic heating are less than 2 TW (Section 9.08.2.7.1), it is clear that there is no difficulty in driving a dynamo at the present day. A heat flow of 7–17 TW also implies an inner-core age of 0.55–1.3 and 0.9–2.3 Gy for models 2 and 1, respectively, assuming a constant core cooling rate. Note that in all cases, an inner core significantly younger than 4.5 Gy is obtained.

Prior to the formation of an inner core, the CMB heat flow had to exceed the adiabatic value Q_k in order to maintain a dynamo for reasons discussed in the preceding text (eqn [12]). For a dynamo requiring an entropy production rate of 200 MW K^{-1}, the core cooling rate had to be roughly 2.5 times as fast to maintain this rate before the onset of inner-core solidification. A geodynamo prior to the onset of inner-core formation is entirely possible but implies either that CMB heat fluxes were higher in the past or that the present-day dynamo is dissipating more heat than it did prior to inner-core formation. Further discussion of these issues is given in Section 9.08.3.2.4.

Figure 2 also shows that, as discussed in the preceding text, a larger CMB heat flow is required for the same entropy production if radioactive heating is important in the present-day Earth. Prior to the existence of the inner core, the effect of radioactive decay on the entropy production is small because the thermodynamic efficiency of radioactive heat production is similar to that of secular cooling (Gubbins et al., 2003; Roberts et al., 2003; Chapter 8.02). Although the presence of potassium also reduces the core cooling rate and thus increases the age of the inner core (eqn [9]), this effect is quite small because the potassium heat production rate is only a small fraction of the likely CMB heat flow. This issue is discussed further in Section 9.08.3.2.5.

In summary, Figure 2 shows that the estimated present-day CMB heat flow of 7–17 TW is consistent with the operation of a dynamo dissipating 0.3–4 TW of heat. Under these

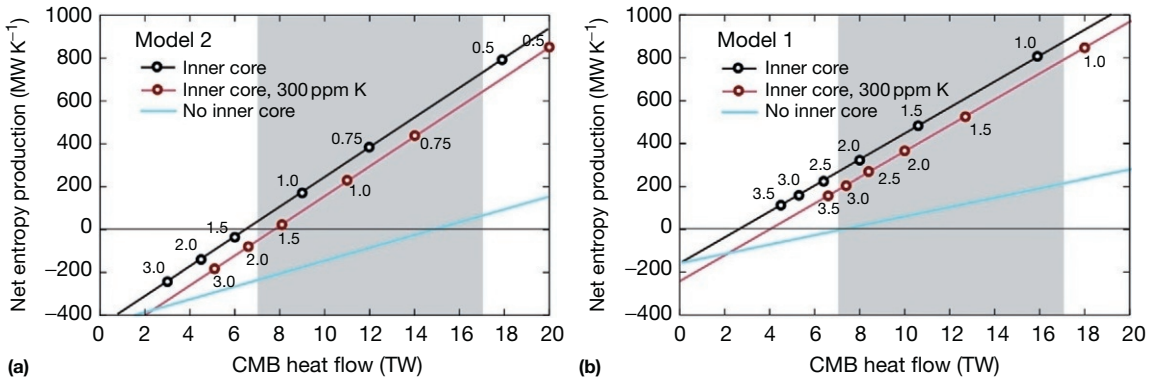


Figure 2 (a) Net entropy production (available to drive the geodynamo) as a function of CMB heat flux, for cases without an inner core, with an inner core, and with an inner core and 300 ppm potassium. Parameters used are for model 2 in Table 2; calculation details are in Chapter 8.02. Labels denote inner-core age in Gy, calculated by assuming a constant core cooling rate and taking $\Delta T_c = 60$ K. (b) The same as for (a), except using the parameters for model 1 in Table 2 (designed to maximize inner-core age) with $\Delta T_c = 150$ K.

circumstances, an inner core could have persisted for 0.55–2.3 Gy if the heat flux stayed constant. Radioactive heating can increase the inner-core age somewhat, but the present-day radioactive heat production is likely only a small fraction of the total energy budget, and thus, the effects are modest. In practice, of course, both the core heat flux and the radiogenic heat production will vary with time; investigating the time evolution of the core and mantle is the subject of the next section.

9.08.3.2.4 Thermal evolution

There are two basic approaches to modeling the thermal evolution of the core. One approach is to start from some assumed initial conditions and evolve the core forward in time, using eqn [9] or its equivalent (Costin and Butler, 2006; Davies, 2007; Mollett, 1984; Nakagawa and Tackley, 2010; Nimmo et al., 2004; Stacey and Loper, 1984; Stevenson et al., 1983; Yukutake, 2000). The initial conditions can be iterated until the correct present-day core parameters (e.g., inner-core size) are obtained and the theoretical geodynamo history compared with the observations. Because the core's evolution depends on the CMB heat flux, such models must simultaneously track the thermal evolution of the mantle. This kind of approach has two principal disadvantages: Firstly, it requires the assumption of initial conditions, which are poorly constrained (**Chapter 9.03**); and secondly, in considering the mantle and the core, the number of important but uncertain parameters (e.g., mantle viscosity) greatly increases.

A second approach is to start from the present-day core conditions and evolve the core backward in time (Buffett, 2002; Buffett et al., 1996; Gomi et al., 2013; Labrosse, 2003). This approach has the advantage of automatically satisfying the present-day observations. However, because diffusion equations are unstable if run backward in time, the evolution of the CMB heat flux cannot be calculated in the same way as it can in the forward models. A common choice is to specify the time evolution of the entropy production in the core, which then specifies both the core cooling rate and the evolution of the CMB heat flux eqn [11]. This approach has the virtue of not requiring any knowledge of the mantle to do the calculations; however, it makes a major assumption in assuming a specific entropy production history for the core. Nonetheless, this approach is both simpler and subject to fewer uncertainties than the alternative and will be focused on here. A major caveat is that the core is assumed to be well mixed throughout; if regions of the core are stably stratified, that will have major (but currently rather unclear) consequences for core thermal evolution. This issue is addressed briefly in **Section 9.08.3.2.7**.

Partly because of geochemical arguments suggesting that the inner core was old (**Section 9.08.3.3.1**), many of the investigations cited in the preceding text have focused on the age of the inner core. While there is a general tendency to find relatively young (~1 Gy) inner cores, the robustness of these results is often unclear because of the large number of poorly constrained parameters, which have to be chosen. Other aims of this section are to tabulate the most important parameters and to investigate the robustness of the thermal evolution results to likely parameter variations. The conclusion that the inner core is young (<2 Ga) is found to be robust, mainly because of the recent increase in estimates of core thermal conductivity k (**Section 9.08.2.2**). An additional conclusion is that the early lower mantle was extensively molten.

9.08.3.2.4.1 Parameters

Of the various parameters discussed in **Section 9.08.2**, we may identify those that will have the largest influence on whether a dynamo can be maintained while producing an ancient core. They are as follows:

1. Thermal conductivity k and thermal expansivity α. A low thermal conductivity or expansivity reduces E_k and thus allows the same rate of entropy production for a lower CMB heat flux (eqn [11]).
2. Gradient of the melting curve. The quantity ΔT_c (eqn [7]) is the change in T_c since the inner core started solidifying and is determined by the relative slopes of the adiabat and the melting curve. A larger ΔT_c results in an older inner core for the same CMB heat flux (or alternatively a higher entropy production rate for an inner core of the same age).
3. The compositional density contrast $\Delta\rho_c$. The larger the value of $\Delta\rho_c$, the higher the entropy production rate for the same rate of cooling.
4. The rate of entropy production required to drive the dynamo. As discussed in **Section 9.08.2.7.1**, this value must exceed 0 MW K^{-1} and is more likely closer to 100 MW K^{-1}.
5. Radioactive heating within the core. Internal heat production reduces the core cooling rate (**Figure 2** and eqn [9]).
6. The extent of stable stratification of the core. A stable layer is likely to reduce the surface field strength and may affect long-term thermal evolution. In the succeeding text, I generally assume that the entire outer core is well mixed (but see **Section 9.08.3.2.7**).

Other factors, such as latent heat, specific heat capacity, and heat of reaction, are either better known than the factors listed in the preceding text or have only a small effect.

Factors 1–3 are known with some uncertainty, while factors 4–6 are less well known. We have therefore adopted two models (**Table 2**) designed to result in maximum and best-guess inner-core ages, respectively. In this way, a conservative assessment may be made of the model variability arising from uncertainties in parameter values.

The calculations shown in the succeeding text take a similar approach to those of Buffett (2002) and Labrosse (2003) and assume a specified rate of entropy production with time. The core temperature is evolved backward from the present-day conditions. The entropy production rate prior to inner-core formation is assumed constant, which allows the CMB heat flux and core cooling rate to be determined. The CMB heat flux during inner-core solidification is assumed to stay constant at the value immediately prior to solidification. The justification for making this assumption is that the CMB heat flux is determined primarily by conditions in the mantle and is thus unlikely to be significantly affected by changing core conditions. This assumption is less reliable for inner cores of greater ages or having larger values of ΔT_c. A result of the assumption is that the entropy production increases significantly when inner-core solidification starts, because of the extra contributions (e.g., latent heat release) to the entropy budget.

Other assumptions could be made. For instance, Labrosse (2003) assumed that the present-day entropy production is some constant factor times the entropy production immediately prior to core formation. It will be shown in the succeeding text that different assumptions of this kind do not significantly affect the results. In theory, one would like to use observations of the Earth's magnetic field to constrain the entropy evolution. For instance, one might anticipate that a higher field strength would lead to greater dissipation and thus higher entropy production. Unfortunately, as discussed in Section 9.08.2.7, neither the observations nor our theoretical understanding of geodynamos is currently good enough to infer how the entropy production has changed. The assumption of constant entropy production prior to inner-core formation at least has the virtue of simplicity; furthermore, since if anything entropy production is likely to have declined with time, this assumption will result in conservatively old inner-core ages.

Figure 3 shows the evolution of various parameters of interest for models 1 and 2 when the net entropy production rate prior to inner-core formation is 100 MW K^{-1}, probably a reasonable value (see Section 9.08.2.7.1). This entropy production rate determines the core cooling rate and thus the heat flux. The present-day heat fluxes are in the range 13–18 TW, in line with expectations (Section 9.08.2.6). The change in CMB heat flow over 4 Gy is modest, a factor 35% or less. Whether such a small change is dynamically plausible is currently unclear and will be discussed further in the succeeding text.

As expected, the heat flux required for model 1 (Figure 3(b)) is lower than for model 2 (Figure 3(a)), because model 1 uses parameter values chosen to favor a long-lived geodynamo. A consequence of this lower heat flux is that the temperature change of the core over 4.5 Gy is smaller than for model 2. The lower heat flux also results in an inner core of greater age, 1.1–1.4 Gy. The present-day rates of entropy production for models 1 and 2 are comparable, because the lower-core growth rate in model 1 is offset by a large compositional density contrast. The total energy released since inner-core formation due to secular cooling, gravity, and latent heat is in the ratio 51:19:30 for model 2 and 70:14:16 for model 1, illustrating the greater relative importance of compositional convection in model 1.

The solid lines and dashed lines in Figure 3 represent cases without and with 300 ppm potassium, respectively. Because of the large present-day CMB heat fluxes involved, adding potassium has little effect on the age of the inner core; however, it does substantially decrease the total change in core temperature (eqn [9]). In fact, in model 1, the core temperature actually initially increases slightly, a probably unphysical result.

Figure 4 shows the same situation as the preceding figure, but now with a net entropy production rate prior to inner-core formation of 0 MW K^{-1}, a lower bound. The lower entropy production results in a reduction in the heat flux required (8–15 TW at the present day) and also a reduction in the amount by which the core has cooled over 4.5 Gy. As a consequence of this reduction in cooling rate, the inner core can persist further back in time. Nonetheless, even adding 300 ppm of potassium, the conservative model 1 still only yields an inner-core age of 1.7 Gy, much less than the age of the Earth.

9.08.3.2.5 Inner-core age and initial core temperature

Figure 5 summarizes the outcomes of several similar models by plotting inner-core age and the initial core temperature against present-day CMB heat flux. As expected, a higher Q_{cmb} results in a younger inner core and a higher initial core temperature. For the same current heat flux, model 1 results in an older inner core than model 2 and also generates a higher rate of entropy production. Models 1 and 2 have very similar initial core temperatures for the same present-day heat flux (as expected), but adding potassium significantly reduces the initial core temperature. Figure 5(b) marks the likely range of lower-mantle solidus and liquidus temperatures, indicating

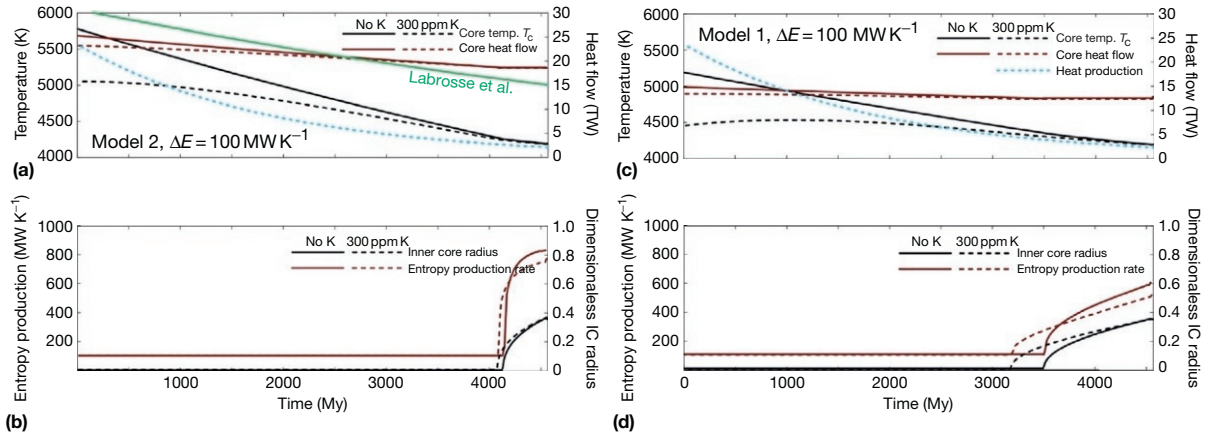


Figure 3 (a) Evolution of core heat flux and core temperature T_c with time for model 2 (Table 2). Heat flux prior to inner-core formation is calculated by fixing entropy production rate at 100 MW K^{-1}. Heat flux after the onset of inner-core formation is kept at the same level as it was immediately prior to solidification. Equations are integrated backward in time from the present-day and $C_r = -10\,100$ m K^{-1} (see Chapter 8.02). Dashed lines represent cases with 300 ppm in core; dotted blue line is heat produced by ^{40}K decay. Green line is model heat flow evolution from Labrosse et al. (2007). (b) Evolution of the entropy production rate and dimensionless inner-core radius R_i/R as a function of time. Note that the entropy production increases when inner-core formation occurs. (c) The same as for (a), except using parameters for model 1 with $C_r = -4900$ m K^{-1}. (d) The same as for (b).

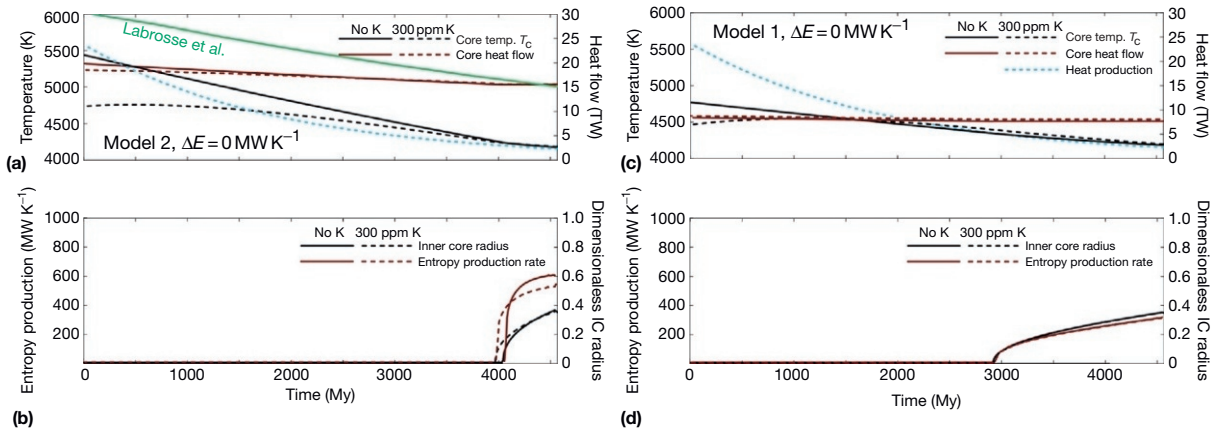


Figure 4 Similar to **Figure 3**, but here with an entropy production rate prior to inner-core formation fixed at 0 MW K^{-1} (i.e., a marginal dynamo).

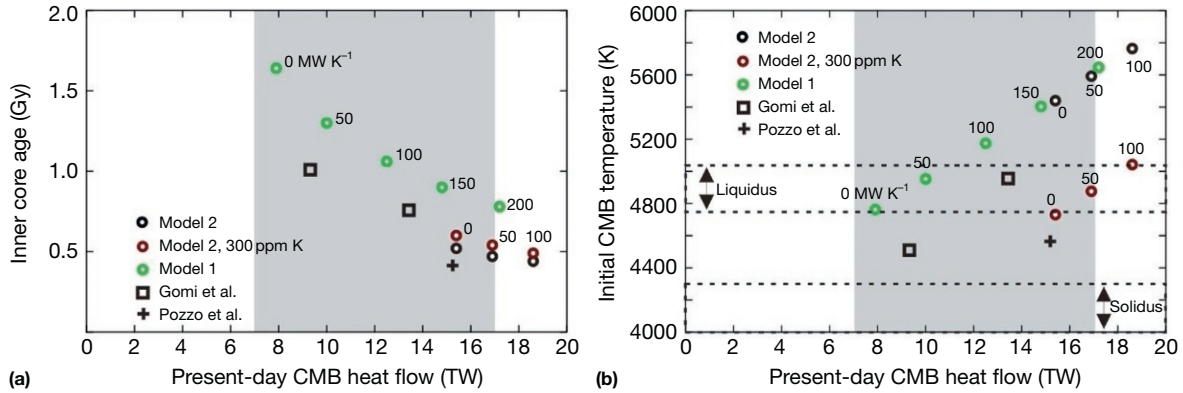


Figure 5 (a) Inner-core age as a function of present-day CMB heat flow; circles are based on a suite of models like those shown in **Figures 3** and **4**. The shaded region indicates the likely present-day heat flow (**Section 9.08.2.6**). Parameters for models 1 and 2 are given in **Table 2**. Labels on individual circles give the fixed entropy production rate (in MW K^{-1}) prior to the onset of inner-core formation. Squares are from Gomi et al. (2013, Figure 4) taking conductivities of 90 and 130 W m^{-1} K^{-1}; cross is from Pozzo et al. (2012, Table 2 case 5). (b) The same as for (a), but plotting initial core temperature against present-day CMB heat flow. The lowermost mantle solidus and liquidus ranges are from Andrault et al. (2011).

that irrespective of the model details, the lower mantle will have initially been partially molten.

A baseline model, with an entropy production rate prior to inner-core formation of 50 MW K^{-1} yields a present-day CMB heat flow of 17 TW, an initial core temperature of 5600 K, and an inner-core age of 0.5 Gy. The present-day entropy production rate is 700 MW K^{-1}, more than enough to drive a dynamo, and the present-day CMB heat flow is marginally compatible with the observational constraints (**Section 9.08.2.6**).

The two key conclusions from **Figure 5** are as follows:

1. The inner core is always young (<1.7 Gy).
2. The initial core temperature significantly exceeded the mantle solidus, so the base of the mantle was initially molten.

Figure 5 also plots comparable results from two recent studies (Gomi et al., 2013; Pozzo et al., 2012). The Gomi et al. (2013) results assume a marginally adiabatic core, permitting somewhat lower CMB heat flows and lower initial core temperatures than our baseline results (model 2). The Pozzo et al. (2012) results are very similar to our baseline model; their present-day entropy production rate of 865 MW K^{-1} is larger than our value of 630 MW K^{-1} because they neglect the heat of reaction term. In any event, **Figure 5** shows that the two main conclusions listed in the preceding text are robust, despite these differences in detail.

It is striking that, at least for the baseline case (model 2), even very modest entropy production rates prior to inner-core formation imply present-day CMB heat flows at the upper end of the likely range. Going further back in time, CMB heat flows increase further (**Figures 3** and **4**) even with the conservative assumption of constant entropy production. This creates two potential problems. First, are such high heat flows compatible with our understanding of mantle behavior? Second, how can such low entropy production rates be reconciled with an ancient dynamo? The first question will be discussed further in the succeeding text (**Section 9.08.3.2.6**). The answer to the second question might simply be that our estimates of the entropy production rates required to drive a dynamo are too high or that the estimated CMB heat flows are too low. Alternatively, additional sources of entropy production (such as exsolution) may have been available.

An apparently young inner core seems inescapable. Although Os isotopic measurements have been advanced in

favor of an old inner core, these observations can be explained by noncore processes (Section 9.08.3.3.1). The onset of inner-core growth might be associated with a change in magnetic behavior, but it is not clear that such a difference would really be detectable in the paleomagnetic record (Aubert et al., 2009; Section 9.08.2.7).

The initial temperature of the core raises further issues. Here, 'initial' refers not to the temperature of the core during accretion (which may have been as high as 10 000 K – see Chapter 9.03), but to the temperature once accretion had finished and the density structure of the Earth resembled the present-day arrangement. Because we are assuming a fixed entropy production rate prior to inner-core growth, we are effectively minimizing the CMB heat flux and thus minimizing the core temperature drop. Hence, the points plotted in Figure 5(b) are lower bounds. Thus, it also appears inescapable that the early mantle was extensively molten (Gomi et al., 2013).

The consequences of this molten state for the CMB heat flux are unclear and in particular depend on whether the melt is more or less dense than the solid. Current thinking favors a dense melt layer (Nomura et al., 2011; Stixrude et al., 2009), although agreement is not universal (Andrault et al., 2012). Assuming a dense melt, the CMB heat flux will ultimately be controlled by the ability of the solid mantle above the melt layer to remove heat. Labrosse et al. (2007) estimated an initial heat flow of 30 TW from the melt layer, declining to 15 TW at the present day. Their present-day flow is shown in Figure 3(a) and resembles our model result; their initial flow is somewhat higher than our value, suggesting that the entropy production rate (and perhaps field intensity?) decreased with time.

9.08.3.2.6 Consequences for the mantle

The aforementioned discussion illustrates an important point: the thermal evolution of the core cannot really be considered separately from the thermal evolution of the mantle. In particular, the evolution of the CMB heat flux controls the thermal evolution of the core and yet is to a large extent determined by mantle processes. Conversely, the early state of the core implies that the lowermost mantle was probably extensively molten, with potentially important consequences for mantle chemistry and dynamics.

A consequence of the requirement of constant entropy production prior to inner-core formation is that the core heat flux does not change very much with time (Figures 3 and 4). Whether a roughly constant heat flux is dynamically plausible remains to be seen, as it depends on the poorly understood details of how the mantle extracts heat from the core. Nonetheless, if the core heat flux has remained roughly constant, there are interesting consequences for the evolution of the mantle.

The evolution of the mantle temperature depends on three factors: internal heat generation, heat added from the core, and the rate of heat loss to the surface. Some petrologic constraints on the evolution of mantle temperatures exist (e.g., Herzberg et al., 2010) and suggest that the mantle potential temperature has changed by perhaps 100–300 K in the last 3.5 Gy (an average secular heat flow of 9 ± 4.5 TW). Given this temperature change, an approximately constant heat flux from the core to the bottom of the mantle, and an assumed mantle heat production rate, the average rate of heat lost by the mantle can be calculated. Assuming a depleted mantle (Davies), the present-day heat production rate is 2.45×10^{-12} W kg^{-1} and the average flow over 3.5 Gy is 16 TW. Taking an average CMB heat flow of 20 TW (Figure 3(a)), the mean mantle heat loss rate is 45 ± 4.5 TW. This is roughly 25% higher than the present-day mantle heat loss rate of 36 TW (Davies) and suggests that (for either linear or exponential decay) the rate of heat loss at 3.5 Ga was roughly 50% higher than the present-day value.

Of course, this calculation is very crude. Nonetheless, it illustrates two important points. First, the secular cooling contribution is small compared to the internal heat production, so whether the depleted upper mantle reflects the overall mantle composition is the key uncertainty. For instance, a deep enriched layer (Boyet and Carlson, 2005) would require much higher mean mantle loss rates. Second, taken at face value, it is surprising that the mantle heat loss rate should change by only a factor of 1.5 over 3.5 Gy. A mantle temperature reduction of 200 K would cause a viscosity increase of about two orders of magnitude. Convective heat flow scales as (viscosity)$^{\beta}$, where $\beta \approx 1/3$ in the absence of complicating factors (Davies). So the predicted change in heat flow is more like a factor of 6 than a factor of 1.5.

There are at least three possible explanations. The first is that the average (bulk) mantle is more enriched in heat-producing elements than the estimate in the preceding text. The second is that the mantle temperature change over 3.5 Gy has been at the lower end of the estimated range, ≈ 100 K. The third is that mantle heat loss, which is controlled by plate tectonics, is less sensitive to temperature than the usual $\beta \approx 1/3$ scaling would suggest. This is not an outlandish hypothesis: Korenaga (2003) has speculated that higher temperatures can even lead to *lower* heat fluxes.

A related question is whether the mantle is truly capable of removing roughly 20 TW from the core. Based on current estimates of mantle thermal conductivity and D″ thickness, it seems improbable that purely conductive heat transport would suffice. However, if the postperovskite layer is weak (Nakagawa and Tackley, 2010) or there is a deep molten layer (Labrosse et al., 2007), heat fluxes in excess of 20 TW appear permissible. This problem is likely to be an area of vigorous future work.

9.08.3.2.7 Stable stratification

There are two ways in which the core might become stably stratified. First, a buoyant compositional layer could accumulate immediately beneath the CMB, formed either by accumulation of light elements expelled from the inner core, from exsolution, or by transfer of light material from the mantle above (see Section 9.08.3.3.1). Such a layer is likely to be stable against penetrative convection from the material beneath (Buffett and Seagle, 2010) and will grow at a rate determined by the supply rate of light elements; however, it may be disrupted by lateral variations in CMB heat flow.

Second, regions of the core in which the heat flux is subadiabatic may also become stably stratified (Loper, 1978), unless disrupted by compositional convection. The location of such a layer depends strongly on the calculation details. For instance, Gubbins et al. (1982), Labrosse et al. (1997), and Lister and Buffett (1998) concluded that a subadiabatic, stratified layer would be found at the top of the core. In the presence

of light elements, this situation is a double-diffusion problem: the effects of compositional and thermal buoyancy are competing against each other (Buffett and Seagle, 2010). As a result, whether such a layer would be stable or not is quite unclear.

Prior to inner-core formation and light element release, the situation is clearer. As shown in Gomi et al. (2013) and Chapter 8.02, the strong variation in thermal conductivity with depth results in a stably stratified layer at the bottom of the core. Penetrative convection from the material above is unlikely to disrupt it (although exsolution within might), and its thickness depends primarily on the CMB heat flow.

Stably stratified layers at the top and bottom of the core are likely to affect thermal and dynamo evolution in different ways. A stagnant conductive layer on top of an active dynamo will screen out short-timescale and short-wavelength fluctuations (e.g., Nakagawa, 2011). A stagnant layer below a dynamo will behave essentially like the solid inner core. A stable bottom layer will not greatly affect the overall thermal evolution of the core, but a stable top layer can potentially act as an insulator (Gubbins et al., 1982; Labrosse et al., 1997), reducing the rate at which the core cools and reducing the entropy available to drive a dynamo.

Figure 6 presents some simplified calculations in which the evolution of a deep stable layer is tracked. The CMB heat flow as a function of time is specified; given this value and the CMB temperature T_c, the location of the top of the stable layer can be calculated using eqns [69] and [81]–[83] of Chapter 8.02. The calculations are truncated at 0.5 Gy BP, since inner-core growth renders them invalid, and the thermal conductivity at the CMB $k_c = 130\ \mathrm{W\ m^{-1}\ K^{-1}}$. Figure 6(a) and 6(c) shows how the stable layer thickness and core temperature evolve when the CMB heat flow is fixed at 20 TW. In this scenario, the unstable layer gets thinner further back in time and in a case without potassium results in a completely stable core prior to 3 Gy BP. Radiogenic heating helps destabilize the core and allows a thin unstable layer to persist over the entire history of the Earth. Figure 6(b) and 6(d) shows similar plots but with a time-variable heat flow similar to that obtained by Labrosse et al. (2007). In this case, the unstable layer gets thicker further back in time, owing to the higher heat flow. Potassium has only a minor effect in this case.

These results are quite sensitive to both the heat flow history and the value of k_c assumed. For instance, taking $k_c = 90\ \mathrm{W\ m^{-1}\ K^{-1}}$ (the value adopted by Gomi et al., 2013) greatly reduces the temporal and radial extent of the stably stratified region. Furthermore, stratified layers can potentially be destabilized by lateral heat flux variations at their boundaries, an effect not included in Figure 6. Nonetheless, these calculations raise the possibility that a significant fraction of the lower core was stably stratified prior to inner-core formation. The dynamo was therefore likely operating only in an outer shell, similar to the situation that pertains today.

9.08.3.3 Compositional Evolution

The compositional evolution of the core since its formation has been a relatively neglected field of study, perhaps because

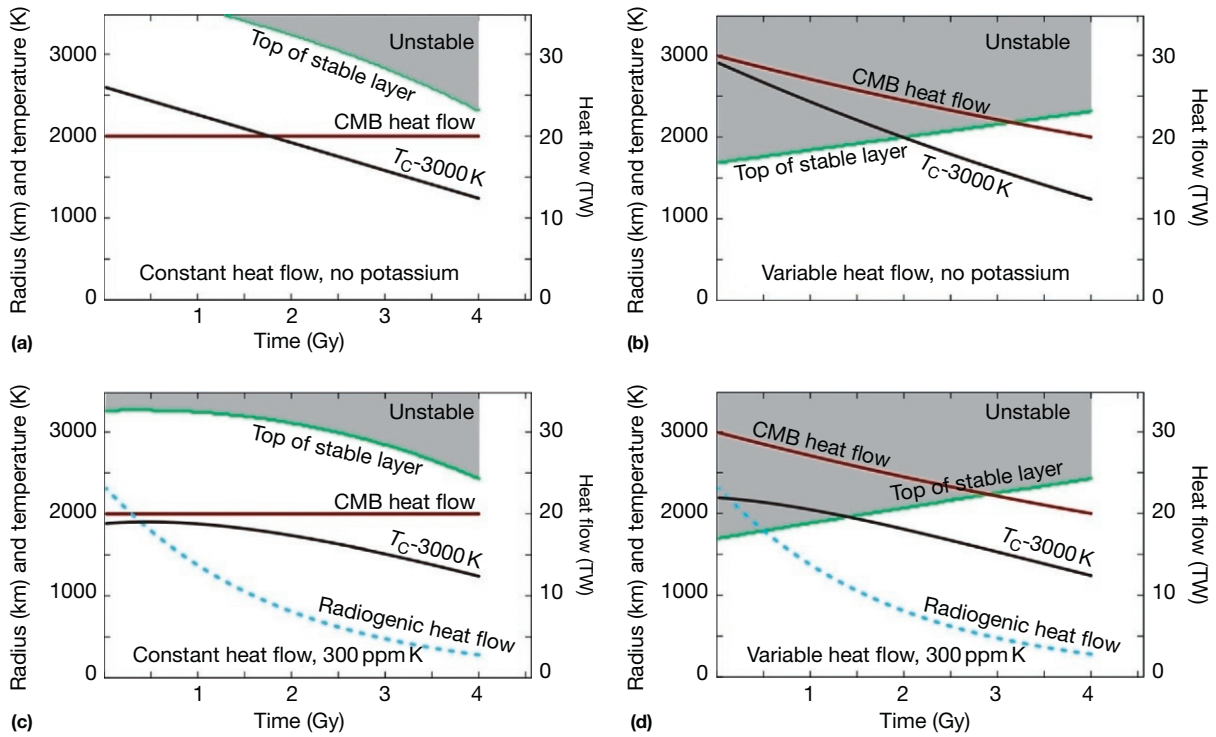


Figure 6 (a) Evolution of stably stratified layer as a function of time for a specified constant CMB heat flow of 20 TW. The location of the top of the stable layer is calculated using eqns [81]–[83] of Chapter 8.02 and taking the thermal conductivity to vary as $k(r) = k_c(1 - (r^2/D_k^2))/(1 - (r_c^2/D_k^2))$ with $k_c = 130\ \mathrm{W\ m^{-1}\ K^{-1}}$, $D_k = 5900$ km, and $r_c = 3480$ km. The evolution of the CMB temperature T_c is also plotted. (b) The same as for (a), but CMB heat flow varies exponentially with time. (c) The same as for (a), but with 300 ppm potassium in the core. (d) The same as for (b), but with 300 ppm potassium in the core.

of the paucity of observational constraints. There are two main ways in which the core composition evolves: through solidification of the inner core and through reaction with the mantle at the CMB. Each of these is dealt with briefly in turn; consequences of inner-core growth and core–mantle interactions are discussed in more detail in Sumita and Bergman and Buffett, respectively.

9.08.3.3.1 Inner-core growth

One potentially observable consequence of inner-core growth is its effect on Os-isotope systematics. The basic hypothesis is relatively simple (Walker et al., 1995): both Re and Pt are presumed to partition preferentially into the outer core relative to Os as solidification proceeds. Since ^{187}Re and ^{190}Pt decay to ^{187}Os and ^{186}Os, respectively, the outer core will become progressively enriched in these Os isotopes relative to stable ^{188}Os. The amount of enrichment depends on the time since inner-core crystallization and the relative partitioning of Re and Pt into the outer core compared to Os. Furthermore, the enrichments in ^{186}Os and ^{187}Os are expected to be coupled if core crystallization occurs (since both are occurring due to the same process). Assuming values for the partition coefficients of Pt, Re, and Os, it has been argued (Brandon et al., 2003; Puchtel et al., 2005) that the onset of inner-core crystallization must have been prior to 3.5 Ga to explain the komatiite Os-isotope ratios.

These results, if correct, are of great significance; unfortunately, they are probably not correct (Meibom, 2008). Other studies have ascribed the osmium anomalies to mantle heterogeneities rather than a core contribution (Baker and Jensen, 2004; Hauri and Hart, 1993; Luguet et al., 2008; Schersten et al., 2004). Moreover, a corresponding signal should also be observed in tungsten and lead isotope anomalies; such signals appear to be lacking in Hawaiian lavas (Lassiter, 2006; Schersten et al., 2004) and Afar plume basalts (Rogers et al., 2010). Finally, measurements of Pt, Re, and Os partition coefficients at intermediate pressures and temperatures give values that are too small to explain the observed isotopic anomalies (Hayashi et al., 2009; Van Orman et al., 2008). Thus, the isotopic argument that the inner core formed prior to 3.5 Ga is not widely accepted and (as seen in the preceding text) is in serious conflict with the geophysical results.

Another consequence of inner-core growth is that it involves the expulsion of one or more light elements into the outer core. Unless they are efficiently mixed into the outer core by convective stirring, these elements will rise to the CMB and generate a stably stratified layer (see Braginsky (2006) and references therein). A similar situation may arise if exsolution of light elements occurs. Rather than the light elements segregating to form a separate layer, it has instead been argued (Buffett and Seagle, 2010) that the addition of elements (specifically Si and O) to the outer core from the mantle drives a chemical reaction, resulting in a silicate-rich layer of light sediments at the top of the core.

The growth of an inner core and associated light element release would presumably disrupt any preexisting stably stratified layer in the lower core (**Figure 6**). In this context, it is perhaps surprising that there is some seismological evidence (see **Chapter 1.23** for reviews) for a stratified layer immediately above the ICB. There is also some evidence for a stably stratified layer at the top of the core. Helffrich and Kaneshima (2010) used seismological observations to infer a 300 km thick subadiabatic layer, though this is not yet generally accepted. Gubbins (2007) used geomagnetic observations to limit the thickness of such a layer to $<$100 km, while Buffett (2010) used Earth nutations to argue that such a layer exists. Future observational constraints on the location and extent of stably stratified layers would be very valuable.

9.08.3.3.2 Core–mantle boundary

The bulk compositions of the core and mantle are strongly out of equilibrium with each other, so the CMB is certainly a region at which reactions will occur (Knittle and Jeanloz, 1991). Oxygen and silicon will be transferred to the outermost core, providing another potential mechanism for stable stratification there (Buffett and Seagle, 2010). Conversely, the lowermost mantle will tend to become enhanced in iron. The thickness of this iron-enriched layer, however, is currently unclear. Solid-state diffusion is too slow to be of interest, while because of the strong negative density contrast, layers generated by capillary action or mantle deviatoric stresses are unlikely to exceed 100 m (Kanda and Stevenson, 2006; Poirier et al., 1998). One possibility is that a so-called morphological instability between Fe and (Mg,Fe)O can result in an Fe-rich layer tens of km thick (Otsuka and Karato, 2012). The layer thickness is of interest because observations based on Earth nutations suggest that the CMB region must include a relatively electrically conductive layer (e.g., Buffett, 2010). If this layer were iron-rich, it would presumably also have a higher thermal conductivity, thus potentially increasing the heat flux across the CMB; further speculation at this point is premature.

9.08.3.4 Summary

Neither the thermal nor the compositional evolution of the core is currently very well understood, because of a lack of observational data (**Section 9.08.2.7**) and considerable uncertainties in the relevant parameter values and the extent of stable stratification. Nonetheless, by assuming that ohmic dissipation was constant prior to inner-core formation and by using models spanning the likely range of parameter values, the thermal history of the core can be investigated, with results summarized in **Figure 5**. Several important points are evident, all of which are robust to likely variations in model parameters.

First, there is no difficulty in maintaining a moderately dissipative dynamo prior to inner-core formation as long as the core is cooling fast enough (cf. **Figures 2–4**); an inner core is not required to drive the early geodynamo.

Second, a higher CMB heat flux implies a more dissipative dynamo and a younger inner core, although the addition of potassium can make the inner core slightly older (**Figure 5(a)**). An estimated present-day CMB heat flux of 7–17 TW yields a present-day entropy production rate of 30–700 MW K^{-1} and an inner-core age of 0.45–1.7 Gy. Prior to inner-core formation, the entropy production rate was probably significantly lower ($<$100 MW K^{-1}).

Third, a dynamo operating for 4 Gy implies initial core temperatures 600–1600 K hotter than that of the present day (**Figure 5(b)**). These temperatures imply that the early lower

mantle was probably extensively molten, likely leading to enhanced CMB heat fluxes (Labrosse et al., 2007).

There are several ways in which future progress in the study of the long-term evolution of the core is likely to be made:

1. Perhaps, the most outstanding question is the extent to which regions of the core were stably stratified. The results in **Figure 6** and Gomi et al. (2013) raise the possibility that a significant fraction of the lower core was stably stratified prior to inner-core formation. The consequences of such a layer for the thermal and magnetic evolution of the core have yet to be thoroughly explored. Inner-core formation (or other processes) may have led to the development of a buoyant layer at the top of the core. Neither the fluid dynamics of this layer nor its implications for the dynamo are thoroughly understood. Observational constraints on stably stratified regions would be very valuable.
2. The likelihood that the lower mantle was initially extensively molten seems inescapable. Although some preliminary work has been carried out (Labrosse et al., 2007; Ulvrova et al., 2012), the consequences for the chemical and thermal evolution of both the mantle and core have not yet been explored in detail.
3. More generally, the evolution of the CMB heat flux is not well constrained, while being absolutely central to the evolution of the core and geodynamo (cf. Davies, 2007; Nimmo et al., 2004). Complicating factors, such as possible melting, the postperovskite phase transition, and perhaps chemical layering, render dynamic models uncertain. The models shown in **Figures 3–4** (which do not include any dynamics) suggest a CMB heat flux, which does not vary greatly over 4 Gy; it is not yet clear whether such results are dynamically plausible. So far, few models have included the available observational constraints on mantle cooling rates, which may help to reduce the possible parameter space (**Section 9.08.3.2.7**).
4. The potential for exsolution of MgO or other species to help drive the dynamo has scarcely been explored but potentially represents an important way of providing enough entropy without requiring excessively high CMB heat flows (**Section 9.08.2.3.1**).
5. Whether or not the core contains any potassium is still an unresolved issue, though the best currently available partitioning experiments argue against it (Nomura et al., 2011). However, the high CMB heat fluxes now preferred mean that the significance of this question has been reduced: potassium has only a small effect on inner-core age (**Figure 5(a)**), though it can still contribute to reducing initial core temperatures (**Figure 5(b)**).
6. Finally, one would expect the changing CMB heat flux, the extent of stable stratification, and the growth of the inner core all to have observable effects on the behavior of the geodynamo. Thus, at least in principle, the paleomagnetic record ought to provide observational constraints on core thermal evolution. The strength of the paleofield is predicted to be insensitive to inner-core formation (Aubert et al., 2009). On the other hand, the time variability of the magnetic field and perhaps the degree of dipolarity are likely to be affected by the inner core; thus, good enough paleomagnetic data may help to tie down when the inner core formed (Coe and Glatzmaier, 2006).

9.08.4 Conclusions

This chapter set out to examine the thermal and compositional evolution of the core, from shortly after its formation to the present day. This period probably involved only two events of importance: the initiation of the geodynamo and the onset of inner-core formation. Prior to inner-core formation, some fraction of the lower core was probably stably stratified; a buoyant stable layer may also have gradually developed at the top of the core, whether from inner-core growth, exsolution, or reaction with the mantle.

Geodynamo activity started at 3.5 Gy BP at the latest; however, it is important to understand that this dynamo could easily have been sustained without an inner core being present. Our baseline model suggests inner-core formation at 0.5 Gy BP. Uncertainty in model parameters could extend the age to 1.7 Gy, but unless a thick, insulating layer exists at the top of the core, the geodynamo must have operated long before inner-core formation.

The change in core temperature over 4 Gy was probably 600–1600 K, implying an early lower mantle that was likely extensively molten. This molten layer was probably responsible for enhanced CMB heat flows, which helped drive the early dynamo.

The present-day core geodynamo is maintained primarily by compositional convection as the inner core solidifies, with exsolution as a possible additional source term. The CMB heat flux is estimated at 12 ± 5 TW and is sufficient to drive a dynamo dissipating 0.1–3.5 TW.

As should be clear, there are several areas that require further study. Firstly, the extent and evolution of core stratification, and its effect on the dynamo, are unclear. Secondly, the evolution of the CMB heat flux over time is currently poorly understood, particularly the effect of lower-mantle melting, and yet has first-order implications for the thermal and chemical histories of both the core and mantle. And finally, future paleomagnetic measurements may help to provide further observational constraints on the evolution of the geodynamo and thus of the core.

Acknowledgments

The author would like to thank Rob Coe for his helpful comments. This chapter was written in January 2013.

References

Abe Y, Ohtani E, Okuchi T, Righter K, and Drake M (2000) Water in the early Earth. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 413–434. Tucson, AZ: University of Arizona Press.

Alfe D, Gillan MJ, and Price GD (1999) The melting curve of iron at the pressures of the Earth's core from ab initio calculations. *Nature* 401: 462–464.

Alfè D, Gillan MJ, and Price GD (2002a) Ab initio chemical potentials of solid and liquid solutions and the chemistry of the Earth's core. *Journal of Chemical Physics* 116: 7127–7136.

Alfè D, Gillan MJ, and Price GD (2003) Thermodynamics from first principles: Temperature and composition of the Earth's core. *Mineralogical Magazine* 67: 113–123.

Alfè D, Price GD, and Gillan MJ (2002b) Iron under Earth's core conditions: Liquid-state thermodynamics and high-pressure melting curve from ab initio calculations. *Physical Review B* 65: 165118.

Alfè D, Price GD, and Gillan MJ (2004) The melting curve of iron from quantum mechanics calculations. *Journal of Physics and Chemistry of Solids* 65: 1573–1580.
Allegre CJ, Poirier J-P, Humler E, and Hofmann AW (1995) The chemical composition of the Earth. *Earth and Planetary Science Letters* 134: 515–526.
Anderson OL (1998) The Gruneisen parameter for iron at outer core conditions and the resulting conductive heat and power in the core. *Physics of the Earth and Planetary Interiors* 109: 179–197.
Andrault D, Bolfan-Casanova N, Lo Nigro G, Bouhifd MA, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259.
Andrault D, Petitgirard S, Lo Nigro G, et al. (2012) Solid–liquid iron partitioning in Earth's deep mantle. *Nature* 487: 354–356.
Aubert J, Labrosse S, and Poitou C (2009) Modelling the palaeo-evolution of the geodynamo. *Geophysical Journal International* 179: 1414–1428.
Backus GE (1975) Gross thermodynamics of heat engines in deep interior of Earth. *Proceedings of the National Academy of Sciences of the United States of America* 72: 1555–1558.
Baker JA and Jensen KK (2004) Coupled 186Os–187Os enrichments in the Earth's mantle–Core–mantle interaction or recycling of ferromanganese crusts and nodules? *Earth and Planetary Science Letters* 220: 277–286.
Belonoshko AB, Ahuja R, and Johansson B (2000) Quasi – Ab initio molecular dynamic study of Fe melting. *Physical Review Letters* 84: 3638–3641.
Biggin AJ, Steinberger B, Aubert J, et al. (2012) Possible links between long-term geomagnetic variations and whole-mantle convection processes. *Nature Geoscience* 8: 526–533.
Bloxham J (2000) Sensitivity of the geomagnetic axial dipole to thermal core–mantle interactions. *Nature* 405: 63–65.
Boehler R (1993) Temperatures in the Earth's core from melting-point measurements of iron at high static pressures. *Nature* 363: 534–536.
Bouhifd MA, Gautron L, Bolfan-Casanova N, et al. (2007) Potassium partitioning into molten alloys at high pressure: Implications for Earth's core. *Physics of the Earth and Planetary Interiors* 160: 22–33.
Boyet M and Carlson RW (2005) Nd-142 evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.
Braginsky SI (1963) Structure of the F layer and reasons for convection in the Earth's core. *Doklady Akademii Nauk SSSR (English Translation)* 149: 1311–1314.
Braginsky SI (2006) Formation of the stratified ocean of the core. *Earth and Planetary Science Letters* 243: 650–656.
Braginsky SI and Roberts PH (1995) Equations governing convection in Earth's core and the geodynamo. *Geophysical and Astrophysical Fluid Dynamics* 79: 1–97.
Brandon AD, Walker RJ, Puchtel IS, Becker H, Humayun M, and Revillon S (2003) 186Os-187Os systematics of Gorgona Island komatiites: Implications for early growth of the inner core. *Earth and Planetary Science Letters* 206: 411–426.
Brown JM and McQueen RG (1986) Phase-transitions, Gruneisen parameter and elasticity for shocked iron between 77 GPa and 400 GPa. *Journal of Geophysical Research* 91: 7485–7494.
Buffett BA (2002) Estimates of heat flow in the deep mantle based on the power requirements for the geodynamo. *Geophysical Research Letters* 29: 1566.
Buffett BA (2010) Chemical stratification at the top of Earth's core: Constraints from observations of nutations. *Earth and Planetary Science Letters* 296: 367–372.
Buffett BA and Christensen UR (2007) Magnetic and viscous coupling at the core-mantle boundary: Inferences from observations of the Earth's nutations. *Geophysical Journal International* 171: 145–152.
Buffett BA, Huppert HE, Lister JR, and Woods AW (1996) On the thermal evolution of the Earth's core. *Journal of Geophysical Research* 101: 7989–8006.
Buffett BA, Matthews PM, and Herring TA (2002) Modeling of nutation and precession: Effects of electromagnetic coupling. *Journal of Geophysical Research* 107: 2070.
Buffett BA and Seagle CT (2010) Stratification of the top of the core due to chemical interactions with the mantle. *Journal of Geophysical Research* 115: B04407.
Bukowinski MST and Akber-Knutson S (2005) The role of theoretical mineral physics in modeling the Earth's interior. In: van der Hilst RD, Bass JD, Matas J, and Trampert J (eds.) *Earth's Deep Mantle: Structure, Composition and Evolution*. *Geophysical Monograph Series*, vol. 160, pp. 165–186. Washington, DC: American Geophysical Union.
Bullard EC (1950) The transfer of heat from the core of the Earth. *Monthly Notices of the Royal Astronomical Society* 6: 36–41.
Cao AM and Romanowicz B (2004) Constraints on density and shear velocity contrasts at the inner core boundary. *Geophysical Journal International* 157: 1146–1151.
Chambers JE (2003) Planet formation. In: *Treatise on Geochemistry*, pp. 461–475. Amsterdam: Elsevier.
Christensen UR (2010) Dynamo scaling laws and applications to the planets. *Space Science Reviews* 152: 565–590.
Christensen UR and Tilgner A (2004) Power requirements of the geodynamo from Ohmic losses in numerical and laboratory dynamos. *Nature* 429: 169–171.
Coe RS and Glatzmaier GA (2006) Symmetry and stability of the geomagnetic field. *Geophysical Research Letters* 33: L21311.
Coltice N and Ricard Y (1999) Geochemical observations and one layer mantle convection. *Earth and Planetary Science Letters* 174: 125–137.
Corgne A, Keshav S, Fei YW, and McDonough WF (2007) How much potassium is in the Earth's core? New insights from partitioning experiments. *Earth and Planetary Science Letters* 256: 567–576.
Costin SO and Butler SL (2006) Modelling the effects of internal heating in the core and lowermost mantle on the Earth's magnetic history. *Physics of the Earth and Planetary Interiors* 157: 55–71.
Davies GF (2007) Mantle regulation of core cooling: A geodynamo without core radioactivity? *Physics of the Earth and Planetary Interiors* 160: 215–229.
De Koker N, Steinle-Neumann G, and Vlcek V (2012) Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T, and heat flux in Earth's core. *Proceedings of the National Academy of Sciences of the United States of America* 109: 4070–4073.
Dunlop DJ and Yu Y (2004) Intensity and polarity of the geomagnetic field during Pre-cambrian time. In: Channell JET, et al. (eds.) *Geophysical Monograph Series*, vol. 145, pp. 85–100. Washington, DC: American Geophysical Union.
Gando A, Gando Y, Ichimura K, et al. (2011) Partial radiogenic heat model for Earth revealed by geoneutrino measurements. *Nature Geoscience* 4: 647–651.
Garnero EJ, Revenaugh J, Williams Q, Lay T, and Kellogg LH (1998) Ultralow velocity zone at the core–mantle boundary. *The Core–Mantle Boundary Region*. *Geodynamics*, vol. 28, pp. 319–334. Washington, DC: American Geophysical Union.
Georg RB, Halliday AN, Schauble EA, and Reynolds BC (2007) Silicon in the Earth's core. *Nature* 447: 1102–1106.
Gessmann CK and Wood BJ (2002) Potassium in the Earth's core? *Earth and Planetary Science Letters* 200: 63–78.
Gessmann CK, Wood BJ, Rubie DC, and Kilburn MR (2001) Solubility of silicon in liquid metal at high pressure: Implications for the composition of the Earth's core. *Earth and Planetary Science Letters* 184: 367–376.
Glatzmaier GA, Coe RS, Hongre L, and Roberts PH (1999) The role of the Earth's mantle in controlling the frequency of geomagnetic reversals. *Nature* 401: 885–890.
Glatzmaier GA and Roberts PH (1996) An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection. *Physica D* 97: 81–94.
Gomi H, Ohta K, Hirose K, et al. (2013) The high conductivity of iron and thermal evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 224: 88–103.
Gubbins D (1977) Energetics of the Earth's core. *Journal of Geophysics* 43: 453–464.
Gubbins D (2007) Geomagnetic constraints on stratification at the top of the Earth's core. *Earth and Planetary Science Letters* 59: 661–664.
Gubbins D, Alfe D, Masters G, Price GD, and Gillan MJ (2003) Can the Earth's dynamo run on heat alone? *Geophysical Journal International* 155: 609–622.
Gubbins D, Alfe D, Masters G, Price GD, and Gillan MJ (2004) Gross thermodynamics of two-component core convection. *Geophysical Journal International* 157: 1407–1414.
Gubbins D, Masters TG, and Jacobs JA (1979) Thermal evolution of the Earth's core. *Geophysical Journal of the Royal Astronomical Society* 59: 57–99.
Gubbins D, Thomson CJ, and Whaler KA (1982) Stable regions in the Earth's liquid core. *Geophysical Journal of the Royal Astronomical Society* 68: 241–251.
Hage H and Muller G (1979) Changes in dimensions, stresses and gravitational energy of the Earth due to crystallization at the inner core boundary under isochemical conditions. *Geophysical Journal of the Royal Astronomical Society* 58: 495–508.
Hauri EH and Hart SR (1993) Re–Os isotope systematics of HIMU and EMII oceanic island basalts from the South Pacific Ocean. *Earth and Planetary Science Letters* 114: 353–371.
Hayashi H, Ohtani E, Terasaki H, and Ito Y (2009) The partitioning of Pt–Re–Os between solid and liquid metal in the Fe–Ni–S system at high pressure: Implications for inner core fractionation. *Geochimica et Cosmochimica Acta* 73: 4836–4842.
Helffrich G and Kaneshima S (2010) Outer-core compositional stratification from observed core wave speed profiles. *Nature* 468: 807–809.
Herzberg C, Condie K, and Korenaga J (2010) Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters* 292: 79–88.
Hewitt J, McKenzie DP, and Weiss NO (1975) Dissipative heating in convective flow. *Journal of Fluid Mechanics* 68: 721–738.
Hillgren VJ, Gessmann CK, and Li J (2000) An experimental perspective on the light element in Earth's core. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 245–264. Tucson, AZ: University of Arizona Press.
Huang HJ, Fei YW, Cai LC, et al. (2011) Evidence for an oxygen-depleted liquid outer core of the Earth. *Nature* 479: 513–516.
Jacobs JA (1998) Variations in the intensity of the Earth's magnetic field. *Surveys in Geophysics* 19: 139–187.

Jephcoat A and Olson P (1987) Is the inner core of the Earth pure iron. *Nature* 325: 332–335.
Kanda RVS and Stevenson DJ (2006) Suction mechanism for iron entrainment into the lower mantle. *Geophysical Research Letters* 33: L02310.
Kleine T, Touboul M, Bourdon B, et al. (2009) Hf–W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochimica et Cosmochimica Acta* 73: 5150–5188.
Knittle E and Jeanloz R (1991) Earth's core–mantle boundary – Results of experiments at high pressures and temperatures. *Science* 251: 1438–1443.
Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophysical Research Letters* 30: 1437.
Labrosse S (2003) Thermal and magnetic evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 127–143.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450: 866–869.
Labrosse S and Macouin M (2003) The inner core and the geodynamo. *Comptes Rendus Geoscience* 335: 37–50.
Labrosse S, Poirier JP, and Le Mouel JL (2001) The age of the inner core. *Earth and Planetary Science Letters* 190: 111–123.
Labrosse S, Poirier JP, and LeMouel JL (1997) On cooling of the Earth's core. *Physics of the Earth and Planetary Interiors* 99: 1–17.
Laio A, Bernard S, Chiarotti GL, Scandolo S, and Tosatti E (2000) Physics of iron at Earth's core conditions. *Science* 287: 1027–1030.
Lassiter JC (2004) Role of recycled oceanic crust in the potassium and argon budget of the Earth: Toward a resolution of the "missing argon" problem. *Geochemistry, Geophysics, Geosystems* 5: Q11012.
Lassiter JC (2006) Constraints on the coupled thermal evolution of the Earth's core and mantle, the age of the inner core, and the origin of the Os-186 Os-188 "core signal" in plume-derived lavas. *Earth and Planetary Science Letters* 250: 306–317.
Lay T, Hernlund J, and Buffett BA (2008) Core–mantle boundary heat flow. *Nature Geoscience* 1: 25–32.
Lay T, Hernlund J, Garnero EJ, and Thorne MS (2006) A post-perovskite lens and D″ heat flux beneath the central Pacific *Science* 314: 1272–1276.
Layer PW, Kroner A, and McWilliams M (1996) An Archean geomagnetic reversal in the Kaap Valley pluton, South Africa. *Science* 273: 943–946.
Lee KKM, Steinle-Neumann G, and Jeanloz R (2004) Ab-initio high-pressure alloying of iron and potassium: Implications for the Earth's core. *Geophysical Research Letters* 31: L11603.
Leng W and Zhong S (2008) Controls on plume heat flux and plume excess temperature. *Journal of Geophysical Research* 113: B04408.
Li J and Fei YW (2003) Experimental constraints on core composition. *Treatise on Geochemistry*, vol. 2, pp. 521–546. Amsterdam: Elsevier.
Lister JR (2003) Expressions for the dissipation driven by convection in the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 145–158.
Lister JR and Buffett BA (1995) The strength and efficiency of thermal and compositional convection in the geodynamo. *Physics of the Earth and Planetary Interiors* 91: 17–30.
Lister JR and Buffett BA (1998) Stratification of the outer core at the core–mantle boundary. *Physics of the Earth and Planetary Interiors* 105: 5–19.
Loper DE (1978) The gravitationally powered dynamo. *Geophysical Journal of the Royal Astronomical Society* 54: 389–404.
Luguet A, Pearson DG, Nowell GM, et al. (2008) Enriched Pt–Re–Os isotope systematics in plume lavas explained by metasomatic sulfides. *Science* 319: 453–456.
Ma YZ, Somayazulu M, Shen GY, Mao HK, Shu JF, and Hemley RJ (2004) In situ X-ray diffraction studies of iron to Earth-core conditions. *Physics of the Earth and Planetary Interiors* 143: 455–467.
Masters G and Gubbins D (2003) On the resolution of density within the Earth. *Physics of the Earth and Planetary Interiors* 140: 159–167.
McDonough WF (2003) Compositional model for the Earth's core. In: *Treatise on Geochemistry*, pp. 517–568. Amsterdam: Elsevier.
McElhinny MW and Senanayake WE (1980) Paleomagnetic evidence for the existence of the geomagnetic field 3.5 Ga ago. *Journal of Geophysical Research* 85: 3523–3528.
Meibom A (2008) Geochemistry – The rise and fall of a great idea. *Science* 319: 418–419.
Mollett S (1984) Thermal and magnetic constraints on the cooling of the Earth. *Geophysical Journal of the Royal Astronomical Society* 76: 653–666.
Morgan JW and Anders E (1980) Chemical composition of the Earth, Venus and Mercury. *Proceedings of the National Academy of Sciences of the United States of America* 77: 6973–6977.
Murthy VR, van Westrenen W, and Fei YW (2003) Experimental evidence that potassium is a substantial radioactive heat source in planetary cores. *Nature* 323: 163–165.
Nakagawa T (2011) Effect of a stably stratified layer near the outer boundary in numerical simulations of a magnetohydrodynamic dynamo in a rotating spherical shell and its implications for Earth's core. *Physics of the Earth and Planetary Interiors* 187: 342–352.
Nakagawa T and Tackley PJ (2004) Effects of thermo-chemical mantle convection on the thermal evolution of the Earth's core. *Earth and Planetary Science Letters* 220: 107–119.
Nakagawa T and Tackley PJ (2010) Influence of initial CMB temperature and other parameters on the thermal evolution of Earth's core resulting from thermochemical spherical mantle convection. *Geochemistry, Geophysics, Geosystems* 11: Q06001.
Nakagawa T and Tackley PJ (2011) Effects of low-viscosity post-perovskite on thermochemical mantle convection in a 3D spherical shell. *Geophysical Research Letters* 38: Q06001.
Nguyen JH and Holmes NC (2004) Melting of iron at the physical conditions of the Earth's core. *Nature* 427: 339–342.
Nimmo F (2007) Energetics of the core. *Treatise on Geophysics*, vol. 8, pp. 31–66. Amsterdam: Elsevier.
Nimmo F, Price GD, Brodholt J, and Gubbins D (2004) The influence of potassium on core and geodynamo evolution. *Geophysical Journal International* 156: 363–376.
Nomura R, Hirose K, Ozawa H, Ohishi Y, and Hirao N (2012) Partitioning of potassium into the Earth's core and implications for thermal history of the Earth. *American Geophysical Union, Fall Meeting, DI33A-2411*.
Nomura R, Ozawa H, Tateno S, et al. (2011) Spin crossover and iron-rich silicate melt in the Earth's deep mantle. *Nature* 473: 199–201.
Okuchi T (1997) Hydrogen partitioning into molten iron at high pressure: Implications for Earth's core. *Science* 278: 1781–1784.
Olson PL, Coe RS, Driscoll PE, Glatzmaier GA, and Roberts PH (2010) Geodynamo reversal frequency and heterogeneous core–mantle boundary heat flow. *Physics of the Earth and Planetary Interiors* 180: 66–79.
Otsuka K and Karato S (2012) Deep penetration of molten iron into the mantle caused by a morphological instability. *Nature* 492: 243–245.
Pavlov V and Gallet Y (2010) Variations in geomagnetic reversal frequency during the Earth's middle age. *Geochemistry, Geophysics, Geosystems* 11: Q01210.
Poirier J-P (1994) Light elements in the Earth's outer core: A critical review. *Physics of the Earth and Planetary Interiors* 85: 319–337.
Poirier J-P, Malavergene V, and Le Mouel JL (1998) Is there a thin electrically conducting layer at the base of the mantle? *The Core–Mantle Boundary Region. Geodynamics*, vol. 28, pp. 131–138. Washington, DC: American Geophysical Union.
Polyakov VB (2009) Equilibrium iron isotope fractionation at core–mantle boundary conditions. *Science* 323: 912–914.
Pozzo M, Davies C, Gubbins D, and Alfe D (2012) Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* 485: 355–359.
Prevot M, Derder ME, McWilliams M, and Thompson J (1990) Intensity of the Earth's magnetic field – Evidence for a Mesozoic dipole low. *Earth and Planetary Science Letters* 97: 129–139.
Puchtel IS, Brandon AD, Humayun M, and Walker RJ (2005) Evidence for the early differentiation of the core from Pt–Re–Os isotope systematics of 2.8-Ga komatiites. *Earth and Planetary Science Letters* 237: 118–134.
Ricolleau A, Fei YW, Corgne A, Siebert J, and Badro J (2011) Oxygen and silicon contents of Earth's core from high pressure metal-silicate partitioning experiments. *Earth and Planetary Science Letters* 310: 409–412.
Righter K and Drake MJ (2003) Partition coefficients at high pressure and temperature. *Treatise on Geochemistry*, vol. 2, pp. 425–449. Amsterdam: Elsevier.
Roberts PH and Glatzmaier GA (2001) The geodynamo, past, present and future. *Geophysical and Astrophysical Fluid Dynamics* 94: 47–84.
Roberts PH, Jones CA, and Calderwood A (2003) Energy fluxes and Ohmic dissipation in the Earth's core. In: Jones CA, et al. (eds.) *Earth's Core and Lower Mantle*. London: Taylor & Francis.
Rogers NW, Davies MK, Parkinson IJ, and Yirgu G (2010) Osmium isotopes and Fe/Mn ratios in Ti-rich picritic basalts from the Ethiopian flood basalt province: No evidence for core contribution to the Afar plume. *Earth and Planetary Science Letters* 296: 413–422.
Rubie DC, Frost DJ, Mann U, et al. (2011) Heterogeneous accretion, composition and core–mantle differentiation of the Earth. *Earth and Planetary Science Letters* 301: 31–42.
Schersten A, Elliott T, Hawkesworth C, and Norman M (2004) Tungsten isotope evidence that mantle plumes contain no contribution from the Earth's core. *Nature* 427: 234–237.
Schonbachler M, Carlson RW, Horan MF, Mock TD, and Hauri EH (2010) Heterogeneous accretion and the moderately volatile element budget of Earth. *Science* 328: 884–887.

Selkin PA and Tauxe L (2000) Long-term variations in palaeointensity. *Philosophical Transactions of the Royal Society of London, Series A* 358: 1065–1088.
Shen GY, Mao HK, Hemley RJ, Duffy TS, and Rivers ML (1998) Melting and crystal structure of iron at high pressures and temperatures. *Geophysical Research Letters* 25: 373–376.
Siebert J, Badro J, Antonangeli D, and Ryerson FJ (2013) Terrestrial accretion under oxidizing conditions. *Science* 339(6124): 1194–1197.
Smirnov AV, Tarduno JA, and Evans DAD (2011) Evolving core conditions ca. 2 billion years ago detected by paleosecular variation. *Physics of the Earth and Planetary Interiors* 187: 225–231.
Stacey FD and Anderson OL (2001) Electrical and thermal conductivities of Fe–Ni–Si alloy under core conditions. *Physics of the Earth and Planetary Interiors* 124: 153–162.
Stacey FD and Loper DE (1984) Thermal histories of the core and mantle. *Physics of the Earth and Planetary Interiors* 36: 99–115.
Stelzer Z and Jackson A (2013) Extracting scaling laws from numerical dynamo models. *Geophysical Journal International* 193: 1265–1276.
Stevenson DJ (2012) How to keep a dynamo running in spite of high thermal conductivity. *American Geophysical Union, Fall Meeting, DI11C-03*.
Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
Stixrude L, De Koker N, Sun N, Mookherjee M, and Karki BB (2009) Thermodynamics of silicate liquids in the deep Earth. *Earth and Planetary Science Letters* 278: 226–232.
Tackley PJ (2012) Dynamics and evolution of the deep mantle resulting from thermal, chemical phase and melting effects. *Earth-Science Reviews* 110: 1–25.
Tarduno JA, Cottrell RD, Watkeys MK, et al. (2010) Geodynamo, solar wind and magnetopause 3.4 to 3.45 billion years ago. *Science* 327: 1238–1240.
Tolstikhin I and Hofmann AW (2005) Early crust on top of the Earth's core. *Physics of the Earth and Planetary Interiors* 148: 109–130.
Ulvrova M, Labrosse S, Coltice N, Raback P, and Tackley PJ (2012) Numerical modelling of convection interacting with a melting and solidification front: Application to the thermal evolution of the basal magma ocean. *Physics of the Earth and Planetary Interiors* 206: 51–66.
Valet JP (2003) Time variations in geomagnetic intensity. *Reviews of Geophysics* 41: 1004.
Van der Hilst RD, De Hoop MV, Wang P, Shim SH, Ma P, and Tenorio L (2007) Seismostratigraphy and thermal structure of Earth's core–mantle boundary region. *Science* 315: 1813–1817.
Van Orman JA, Keshav S, and Fei YW (2008) High pressure solid/liquid partitioning of Os, Re and Pt in the Fe–S system. *Earth and Planetary Science Letters* 274: 250–257.
Verhoogen J (1961) Heat balance of the Earth's core. *Geophysical Journal of the Royal Astronomical Society* 4: 276–281.
Walker RJ, Morgan JW, and Horan MF (1995) Os-187 enrichment in some plumes: Evidence for core–mantle interaction? *Science* 269: 819–822.
Wanke H and Dreibus G (1988) Chemical composition and accretion history of terrestrial planets. *Philosophical Transactions of the Royal Society of London* 325: 545–557.
Wood BJ and Halliday AN (2005) Cooling of the Earth and core formation after the giant impact. *Nature* 437: 1345–1348.
Wu B, Driscoll P, and Olson P (2011) A statistical boundary layer model for the mantle D″ region *Journal of Geophysical Research* 116: B12112.
Yoo CS, Holmes NC, Ross M, Webb DJ, and Pike C (1993) Shock temperatures and melting of iron at Earth core conditions. *Physical Review Letters* 70: 3931–3934.
Yukutake T (2000) The inner core and the surface heat flow as clues to estimating the initial temperature of the Earth's core. *Physics of the Earth and Planetary Interiors* 121: 103–137.

This page intentionally left blank

9.09 The History of the Earth's Rotation: Impacts of Deep Earth Physics and Surface Climate Variability

WR Peltier, University of Toronto, Toronto, ON, Canada

© 2015 Elsevier B.V. All rights reserved.

9.09.1	**Polar Motion and Length-of-Day Variations Through (Geologic) Time**	221
9.09.2	**Theoretical and Observational Background I: Angular Momentum Conservation on Subannual to Interannual Timescales**	222
9.09.2.1	Theoretical Preliminaries	222
9.09.2.2	Variations in the l.o.d.	223
9.09.2.3	Polar Motion: The Annual and Chandler Wobbles	225
9.09.3	**Theoretical Background II: The Viscoelastic Rotational Response to the Late Pleistocene Ice-Age Cycle**	227
9.09.3.1	The Global Theory of the GIA Process	227
9.09.3.2	Computation of the Rotational Response to Ice–Earth–Ocean Interactions in the Ice Age	229
9.09.3.3	The 'Equivalent Earth Model' Approach of : An Approximate Solution	231
9.09.3.4	A Formulation that Retains the Difference Between $k_2^T(s=0)$ and k_f	231
9.09.3.5	Models of the History of Variations in the Elements of Earth's Moment of Inertia Tensor	233
9.09.4	**Observations of Millennial Scale Secular Variations in the Earth's Rotation Anomalies and the Impact of the Global Warming Era upon Them**	237
9.09.4.1	Ancient Eclipse Observations and the Nontidal Acceleration of Rotation	238
9.09.4.2	Millenial Timescale Polar Wander and the Glaciation–Deglaciation Cycle	239
9.09.4.3	Global Warming Era Shifts in the Millennial Timescale Trends in the Earth's Rotation Anomalies Based upon Modern Earth Orientation Data Sets	241
9.09.5	**Earth's Rotational Response to the Cyclic Glaciation Cycle of Late Pleistocene Time: Data–Model Comparisons**	243
9.09.5.1	A Database of Holocene Relative Sea-Level Histories	244
9.09.5.2	The Influence of Rotational Feedback upon Postglacial Relative Sea-Level History and Its Impact upon Predictions of the Earth's Rotation Anomalies	244
9.09.5.3	Measurements of the Expected Quadrupolar 'Signature' of Rotational Feedback in Postglacial Sea-Level Histories	247
9.09.6	**Grace Satellite Inferences of Geoid Height Time Dependence: The Combined Influences of Ice Age and Modern Land-Ice Melting**	255
9.09.6.1	The Geoid Equation: Analytic Results for the Stokes Coefficients of Degree 2 and Order 1	256
9.09.6.2	The Exact Geoid and the Approximation to It Observable by GRACE	258
9.09.6.2.1	The load contribution	259
9.09.6.2.2	The rotational contribution	259
9.09.6.2.3	The total degree 2 and order 1 Stokes coefficients	260
9.09.6.3	GRACE Estimates of the Modern Rate of Land-Ice Melting	261
9.09.7	**The Impact of Variations in the Geometry of the Earth's Orbit Around the Sun upon Climate System Evolution**	264
9.09.7.1	The Astronomical Imprint on Oxygen-Isotopic Records from Deep-Sea Sedimentary Cores	265
9.09.7.2	'Orbital Tuning' and the Age of the Brunhes–Matuyama Geomagnetic Polarity Transition	265
9.09.8	**The Earth's Rotation Variations, Mantle Convective Mixing, and Plate Tectonics-Related Processes**	267
9.09.8.1	Tidal Despinning of the Earth's Rate of Axial Rotation: Tidal Rhythmites and the l.o.d.	268
9.09.8.2	Polar Wander and the Impact of Excess Oblateness due to Mantle Convection	271
References		275

9.09.1 Polar Motion and Length-of-Day Variations Through (Geologic) Time

Observations of the evolving state of planetary rotation reveal the occurrence of variability on timescales ranging from daily, to annual, to interannual, decadal, and millennial and extending even to the timescale of hundreds of millions of years on which the process of mantle convection governs planetary evolution and to the 4.56 billion-year age of the planet itself. These variations in the state of rotation are most usefully discussed in terms of changes in the rate of planetary rotation about the instantaneous spin axis and thus variations in the length of the day, on the one hand, or in terms of the wobble of the spin axis as observed in a body-fixed frame of reference on the other. On relatively short subannual to annual timescales, variations in the length of day have been clearly shown to arise

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00161-5

primarily as a consequence of the exchange of angular momentum between the solid Earth and its overlying atmosphere (e.g., Hide et al., 1980) and oceans. On the interannual timescale, an important recent discovery concerning length-of-day (l.o.d.) variability has been the documentation of a significant excitation associated with El Niño Southern Oscillation (ENSO) events (Cox and Chao, 2002; Dickey et al., 2002). Concerning the sources of wobble excitation, it is clear that on the timescale of the seasonal cycle of climate change, the excitation of the annual component of wobble variability is due to the interhemispheric exchange of atmospheric mass. The Chandler wobble, however, a free oscillation of the Earth's spin axis in a body-fixed frame of reference with a period close to 14 months, is apparently significantly forced by the dynamic state of the oceans (e.g., Gross, 2000) and by the atmosphere. These most recent analyses of the problem of Chandler wobble excitation, to be reviewed in what follows, appear to have finally resolved what had remained an unresolved problem for decades.

On the longer timescale of millennia, both of these 'anomalies' in the Earth's rotation exhibit apparent secular variations that are caused primarily by the Late Pleistocene cycle of glaciation and deglaciation (Peltier, 1982) that has been an enduring feature of climate system variability for the past 900 000 years of Earth history (e.g., Deblonde and Peltier, 1991). A primary focus of the discussion to follow in this chapter will be upon the manner in which, through the process of glacial isostatic adjustment (GIA), these ice-age engendered variations in the Earth's rotation feed back upon postglacial relative sea-level history, thus enabling detailed tests to be performed on the quality of the theory that has been developed to compute the rotational response to the GIA process (Peltier, 2005). The importance of an accurate attribution of the source of excitation of the observed secular changes in the l.o.d. and polar motion to the GIA process concerns the important role that these observations may be invoked to play in the inference of the viscosity of the deep Earth, a parameter that is required in the construction of models of the mantle convection and continental drift processes.

An interesting additional aspect of the history of the Earth's rotation on the timescale of the Late Pleistocene ice-age cycle concerns the way in which temporal variations in the precession and obliquity components of the evolving geometry of the Earth's orbit around the Sun, forced by gravitational n-body effects in the solar system, have been employed to refine the timing of the ice-age cycle itself (Shackleton et al., 1990).

On the very longest timescales on which the thermal evolution of the planet is governed by the mantle convection process, there also exists the distinct possibility that relatively rapid and large amplitude changes in the rotational state could have occurred in association with an 'avalanche effect' during which the style of the mantle convective circulation switches from one characterized by significant radial layering of the thermally forced flow to one of 'whole mantle' form (e.g., Solheim and Peltier, 1994a,b). This process could conceivably act so as to induce the inertial interchange true polar wander (IITPW) instability that was suggested initially by Gold (1955) and that has recently been invoked by Kirschvink et al. (1997) as plausibly having occurred in the early Cambrian period of Earth history.

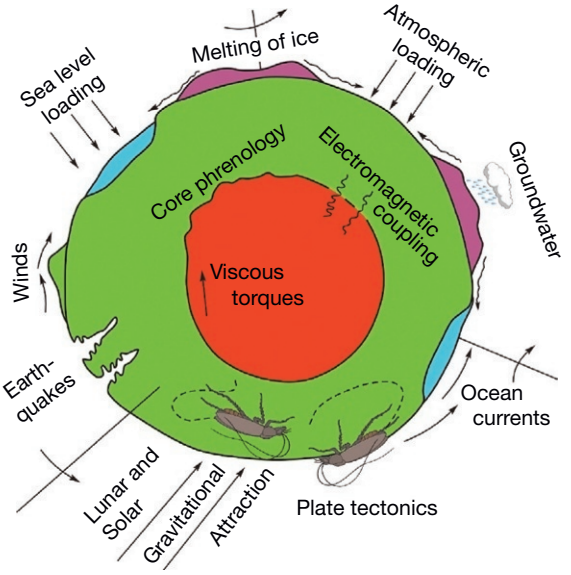


Figure 1 Schematic diagram illustrating the range of processes that contribute to the excitation of variations in the rotational state of the planet. The centipedes in the figure, which are a modification of the beetles employed by Gold (1955) for the same illustrative purpose, are intended to represent the contribution to rotational excitation due to the mantle convection process. This schematic is a modification of that in Lambeck (1980a) paper commenting upon the important paper of Hide et al. (1980).

Figure 1 provides a schematic depiction of the extremely broad range of processes that contribute to the excitation of variations in the Earth's rotation on all timescales. The centipedes in the sketch, following the colorful analogy by Gold (1955), are intended to represent, by their ability to move over the surface and thereby slowly (?) modify the moment of inertia tensor of the planet, the excitation of a rotational response due to the action of the internal mantle convective mixing process. How this component of the rotational excitation might be most accurately described is still a matter of considerable debate, as will be discussed in the final section of this chapter.

9.09.2 Theoretical and Observational Background I: Angular Momentum Conservation on Subannual to Interannual Timescales

9.09.2.1 Theoretical Preliminaries

An instantaneous state of the Earth's rotation is described entirely by the three Cartesian components of the planet's evolving angular velocity vector with components $\omega_i(t)$. The evolution of the angular velocity vector is itself determined by a solution of the classical Euler equation describing the conservation of angular momentum of a system subjected to external torques τ_j, as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(J_{ij}\omega_i\right) + \in_{ijk}\omega_j J_{kl}\omega_l = \tau_j \qquad [1]$$

in which J_{ij} are the elements of the moment of inertia tensor and $\in_{ijk}$ is the Levi-Civita (alternating) tensor. Restricting

attention to small departures from a modern state of steady rotation with angular velocity Ω_o, we may construct a solution to eqn [1], accurate to the first order in perturbation theory, by expanding

$$\omega_i = \Omega_o(\delta_{i3} + m_i);\ m_i = \omega_i/\Omega_o \quad [2a]$$

$$J_{11} = A + I_{11} \quad [2b]$$

$$J_{22} = B + I_{22} \quad [2c]$$

$$J_{33} = C + I_{33} \quad [2d]$$

$$J_{ij} = I_{ij},\ i \neq j \quad [2e]$$

Substitution of these expansions into eqn [1], keeping only terms of first order, leads to the standard set of governing equations for polar wander and the l.o.d., respectively (see, e.g., Munk and MacDonald (1960)), as

$$\left.\begin{aligned} \frac{dm_1}{dt} + \frac{(C-B)}{A}\Omega_o m_2 &= \Psi_1 \\ \frac{dm_2}{dt} + \frac{(C-A)}{B}\Omega_o m_1 &= \Psi_2 \end{aligned}\right\} \text{ polar wander} \quad [3a,b]$$

$$\left.\frac{dm_3}{dt} = \Psi_3\right\} \text{ length of day} \quad [3c]$$

in which the 'excitation functions' are defined as

$$\Psi_1 = \left(\frac{\Omega_o}{A}\right) I_{23} - \frac{(dI_{13}/dt)}{A} + \tau_1 \quad [4a]$$

$$\Psi_2 = -\left(\frac{\Omega_o}{B}\right) I_{13} - \frac{(dI_{23}/dt)}{B} + \tau_2 \quad [4b]$$

$$\Psi_3 = -\left(\frac{I_{33}}{C}\right) + \tau_3 \quad [4c]$$

Now, it is important to recognize that we are here distinguishing between two different ways of describing the rotational excitation, due, respectively, to externally applied torques τ_i and to 'internally' originating perturbations I_{ij} to the moment of inertia tensor. In this section, a brief review will be provided of our current understanding of the nature and origins of the variations in the Earth's rotation that occur on subannual to interannual timescales. These variations are most simply understood by considering the rotational response of an 'almost rigid' solid Earth to externally applied torques derivative of the action upon it of the atmosphere and oceans and to the redistributions of mass that occur both subseasonally and interannually within these fluid components of the Earth system. As we will see in Section 9.09.4, it will be more natural to describe the process involved in the excitation of millennial timescale variations of the Earth's rotation by invoking 'internally' generated changes in the moment of inertia tensor.

9.09.2.2 Variations in the l.o.d.

Focusing first upon changes in the l.o.d. on relatively short timescales, the connection of variations on these timescales to the exchange of angular momentum between the solid Earth and overlying atmosphere was first established most clearly in the important paper by Hide et al. (1980). It will be useful to revisit their initial order of magnitude arguments here in order to fix ideas. If one denotes the total axial angular momentum of the Earth system by the sum $M + M_a$, as in Hide et al., which has the approximate value $5.85 \times 10^{33}\,\mathrm{kg\,m^2\,s^{-1}}$ and divides this by the axial moment of inertia that the atmosphere plus solid Earth system would have if the whole system were in a state of rigid body rotation, namely, $I + I_a \approx 8.04 \times 10^{37}\,\mathrm{kg\,m^2}$, then one would obtain for the angular velocity of the Earth system $\Omega_R \approx 0.726 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$. A consequence of the fact that the atmosphere of the Earth is differentially heated by the Sun is that the complex general circulation of its atmosphere is characterized by a marked differential rotation between it and the solid Earth. This leads to a net increase in the angular momentum of the atmosphere by the (approximate) amount $\delta M_a \approx 1.5 \times 10^{26}\,\mathrm{kg\,m^2\,s^{-1}}$ as determined on the basis of direct atmospheric observations. Given the approximate value of the moment of inertia of the atmosphere, $I_a \approx 1.42 \times 10^{32}\,\mathrm{kg\,m^2}$, this leads to an approximate value for the change in angular velocity of magnitude $\delta M_a / I_a \approx 10^{-6}\,\mathrm{rad\,s^{-1}} \approx 3 \times 10^{-8}\Omega_R$, or given that the l.o.d. is just $2\pi/\Omega_R$, one obtains an approximate magnitude for the change in the length of the day due to the solar differential heating forced general circulation of the atmosphere of 3×10^{-3} s. Detailed astronomical measurements by the Bureau Internationale de l'Heure (BIH) directly demonstrate that on the annual and interannual timescales, this is in fact the order of magnitude of the variations in the l.o.d. observed.

Figure 2 reproduces a figure from the original paper of Hide et al. in which a detailed analysis of the angular momentum budget of the atmosphere was performed by evaluating the axial component of the relative angular momentum m_a as

$$m_a = \frac{a^2}{g}\int_{-\pi/2}^{+\pi/2}\int_0^{2\pi}\int_0^{p_g} u(p, \theta, \lambda, t)\cos^2\theta\, dp\, d\lambda\, d\theta \quad [5]$$

in which a is the radius of the Earth, g is the surface gravitational acceleration, u is the zonal component of the velocity of atmospheric air in a frame of reference rotating with the solid Earth, p is atmospheric pressure, p_g is the pressure at the surface of the Earth, θ is the latitude, λ is the longitude, and t is the time. The radial part of the required volume integral has been converted to an integral over pressure by employing the hydrostatic approximation. Based upon the use of independent sets of reanalysis data from the UK Meteorological Office and from the US National Meteorological Center, the authors inferred the changes in the l.o.d. that should have occurred if the atmosphere was simply exchanging its axial angular momentum with that of the solid Earth. These estimates of temporal changes in the l.o.d. were compared directly to the astronomically derived measurements provided by the BIH over a 9-month period extending from the winter of 1978 to the summer of 1979. These results, which have since only been reconfirmed by later and more detailed analysis, clearly establish that on these short timescales, changes in the l.o.d. are primarily associated with the exchange of angular momentum between the solid Earth plus oceans and its overlying atmosphere.

Considered over a period of many decades, rather than years, one begins to see additional structure in the variability in the l.o.d. that includes the measurable influence of an effect due to the changing axial component of the moment of inertia of the solid Earth, an influence that is generally measured in

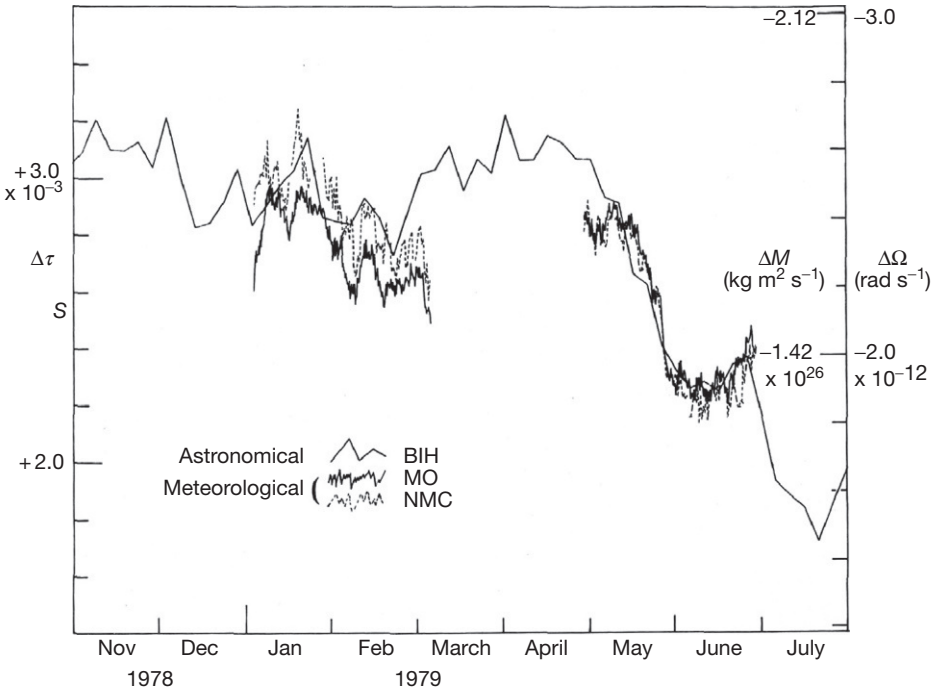


Figure 2 Comparison of the expected changes in the length of the day, represented by the parameter $\triangle\tau$ measured in milliseconds and derived from the meteorological results (either UK Meteorological Office (MO) or the US National Meteorological Center), with the observed changes based upon the astronomical measurements of the Bureau International de l'Heure (BIH). Reproduced from Hide R, Birch NT, Morrison LV, Shea DJ, and White AA (1980) Atmospheric angular momentum fluctuations an changes in the length of the day. *Nature* 286: 114–117.

terms of a parameter $\dot{J}_2$, which will be discussed at length in subsequent sections of this chapter (the 'dot' above the symbol 'J' here represents time differentiation). The parameter J_2 is itself the nondimensional degree 2 and order 0 coefficient in the spherical harmonic expansion of the gravitational field of the planet. The connection between this parameter and the angular velocity of the planet and thus the l.o.d. is given by the following simple expression:

$$\dot{J}_2 = \frac{3}{2}\frac{C}{a^2 m_e}\dot{m}_3 \qquad [6]$$

in which 'a' and 'm_e' are the Earth's radius and mass, respectively, and $m_3 = \omega_3/\Omega_0$.

Superimposed upon the secular trend in J_2, an effect we will show to be unambiguously connected to the GIA process, the analyses by Cox and Chao (2002) were the first in which an interannual timescale variation of this quantity was detected that was apparently associated with the ENSO phenomenon. This internal oscillatory ENSO mode of climate system variability is supported by coupling between the atmosphere and the oceans across the equatorial Pacific region and recurs on a timescale that varies from 3 to 7 years. Further analysis of this interannual variability of J_2, which at the time of writing is still not entirely understood, was presented by Dickey et al. (2002) in which an attempt was made to ascribe the El Niño influence to a particular related process having to do with "a recent surge in subpolar glacial melting and by mass shifts in the Southern, Pacific and Indian Oceans." Figure 3, from the Dickey et al. paper, displays a set of time series of the parameter J_2 that marked (**A**) being the time series of the raw data from the BIH. Although prior to approximately late 1996 the time series reveals a consistent secular decrease in the value of this parameter, such that a least-squares best fit of a straight line to the data delivers the best estimate of $\dot{J}_2$ of approximately -2.9×10^{-11} year^{-1}, beginning in late 1996, the onset of a significant deviation developed from this secular trend. The timing of this excursion in the time series was interpreted as being associated with the onset of the intense El Niño event that occurred in 1998. As we will discuss in Section 9.09.4.3 in the succeeding text, this interpretation has become questionable. The remaining time series shown in Figure 3 display the alterations to it that obtain upon successive removal of the influence of the forcing associated with four plausible influences, respectively, in (**B**) integrated ocean effects as determined using a data-constrained ocean model, in (**C**) subpolar glacier wasting effects, in (**D**) integrated atmospheric effects derived on the basis of National Center for Environmental Prediction (NCEP) reanalysis data, and in (**E**) integrated groundwater effects based upon NCEP reanalysis data. Although it remains an issue as to whether the attribution of the influence of ENSO upon the J_2 excursion as due to an enhanced wasting of subpolar glaciers is correct, there can be no doubt that the warming associated with the strong 1998 ENSO event appears to have induced a substantial variation in the l.o.d. Although the BIH data prior to 1984 were not of sufficient accuracy to generate a longer time series of the variability in this parameter that is equally useful as that shown in Figure 3, it is nevertheless extremely important that the dominant secular variation in the time series is essentially the same as that previously inferred on the basis of the analysis of ancient eclipse observations as we will discuss in the following sections of this chapter.

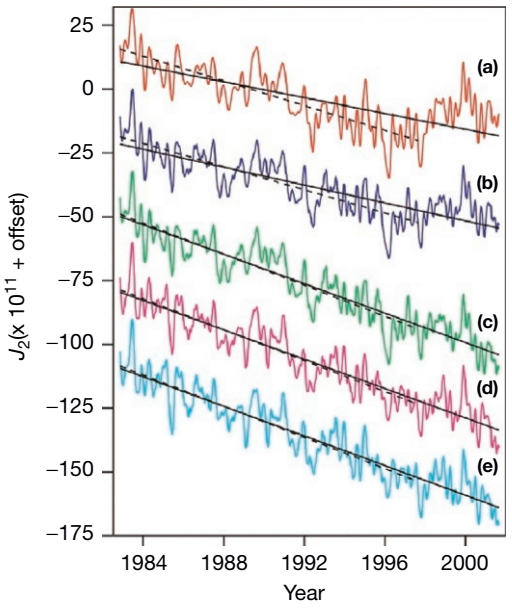


Figure 3 (a) The geodetically observed time series of the parameter J_2 over the time period 1983–2002 and residuals (b)–(e) obtained by successive removal of modeled source terms related, respectively, to the processes of (b) integrated ocean effects, (c) subpolar glacial effects, (d) integrated atmospheric effects, and (e) integrated groundwater effects. Note that these analyses did not anticipate the results of Roy and Peltier (2011), which demonstrated that the rate of secular variation of this parameter underwent a marked change early in the 1990s (see **Section 9.09.4.3**). Reproduced from Dickey JO, Marcus SL, de Viron O, and Fukumori I (2002) Recent Earth oblateness variations: Unravelling climate and postglacial rebound effects. *Science* 298: 1975–1977.

9.09.2.3 Polar Motion: The Annual and Chandler Wobbles

The Chandler wobble or free Eulerian nutation of the spin axis of the Earth in a body-fixed frame of reference was first discovered by Chandler (1891). A detailed theory of the Chandler wobble from a 'normal-mode' perspective was first presented by Dahlen and Smith (1975) and has been more fully developed since by Wahr (1982, 1983) whose investigation focused upon the important influence of the atmosphere and oceans on the phenomenon. Although manipulation of eqn [3a,b] in the preceding text predicts the existence of a simple harmonic 'free' oscillation of frequency $\sigma_R = \Omega_0(C-A/A)$, to which I will subsequently refer as the Chandler wobble frequency of the rigid Earth, this expression substantially errs in its prediction of the period of the actual Chandler wobble because of the importance of these components of the Earth system and the existence of the fluid core upon wobble dynamics.

Furthermore, this system of equations would predict a Chandler wobble that would have infinite 'Q' in the sense that, once excited, it would never decay in amplitude.

Aside from the Chandler wobble, which has a variable amplitude that ranges between ~100 and 200 milliarcseconds (mas) in the angular displacement of the pole, several additional physical processes contribute to the complete range of polar motion that is observed using the space-geodetic techniques of very long-baseline interferometry (VLBI), satellite and lunar laser ranging, and Global Positioning System interferometry. These include (1) a forced annual wobble with an amplitude of ~100 mas, which occurs in response to the seasonal variability of the atmosphere and oceans, (2) a weaker but quasiperiodic variation on decadal timescales with an amplitude of ~30 mas that is often referred to as the Markowitz wobble, and, the most important from the perspective of this chapter, (3) a linear trend corresponding to a secular drift in the position of the pole that is presently occurring at a rate of ~3.5 mas year^{-1}. The latter ultra-low-frequency feature, which will be a primary focus in the main body of this chapter to follow, is the polar motion counterpart to the secular variation in the l.o.d. represented in terms of $\dot{J}_2$ and discussed in the last subsection. The rate of 3.5 mas year^{-1} corresponds to a rate of drift of the pole across the surface of ~0.95° My^{-1}.

Although each of the highest-frequency contributions to the motion of the pole has been studied in depth, it has been the development of a detailed understanding of the Chandler wobble that has attracted the greatest interest, as the explanation of its excitation mechanism has proved to be extremely elusive. Attempts were made to ascribe its excitation to the atmosphere (Wahr, 1983; Wilson and Haubrich, 1976), to core–mantle interactions (Jault and Le Mouel, 1993; Rochester and Smiley, 1965), to the occurrence of earthquakes (Gross, 1986; Sauriau and Cazanave, 1985), and even to continental water storage (Chao et al., 1987; Kuehne and Wilson, 1991). More recent research has however suggested that the primary excitation may in fact be oceanographic (Gross, 2000), although aided to significant degree by the atmosphere.

In attempting to attribute the excitation of the Chandler wobble to any particular forcing function, it is also important to recognize that this important mode of rotational free oscillation also has a finite 'Q' so that if all sources of excitation were to vanish, the amplitude of the polar motion would decay to zero. For this reason, a simple heuristic model, derivable from the linearized Liouville equations [3a,b] in the preceding text, given the extremely small amplitude of the observed motion, is just (e.g., Lambeck, 1980b)

$$\Gamma + \frac{i}{\sigma_{CW}}\frac{d\Gamma}{dt} = \chi(t) \qquad [7]$$

in which $\Gamma = x_p - iy_p$ where x_p and y_p are, respectively, the displacements of the pole from the Conventional International Origin (CIO) along the Greenwich and 90°W meridians and σ_{CW} is the complex Chandler wobble frequency of the real anelastic Earth for which the 'Q' of the Chandler wobble has been estimated by Wilson and Vincente (1990) to be 179 with 1σ bounds of 74 and 789. Equation [7] describes simple harmonic motion in the complex plane subject to the forcing represented by the function $\chi(t)$, an excitation function that Wahr (1982) had shown may be written as

$$\chi(t) = \frac{1.61}{\Omega_o(C-A)}\left[h(t) + \frac{\Omega_o I(t)}{1.44}\right] \qquad [8]$$

in which the complex valued function $I(t) = I_{13}(t) + iI_{23}(t)$ where I_{13} and I_{23} are the perturbations of the inertia tensor associated with the variations in the mass distributions in the atmosphere and the oceans that may be responsible for wobble excitation. The function $h(t) = h_1(t) + ih_2(t)$ represents changes

of relative angular momentum associated, for example, with atmospheric winds and ocean currents. The numerical factors 1.61 and 1.44, respectively, account for the influences of core decoupling and the yielding of the solid Earth under the weight of the changing surface mass load.

In the paper by Gross (2000), detailed analyses were presented of Chandler wobble excitation that were based upon the assumption of the validity of the estimates of the Chandler period and Q by Wilson and Vincente (1990), an analysis in which the estimates were based upon the longest time series of polar motion data available at that time (86 years), the only study in which the duration of the data set analyzed was longer than the e-folding time for the decay of the Chandler amplitude. Since the polar motion is resonant at the Chandler period, most analyses of the excitation mechanism have been conducted in the frequency domain by direct comparison of the spectrum of the rotational response to the spectrum of the excitation as computed on the basis of assumed *a priori* knowledge of a range of plausible excitation mechanisms. **Figure 4(a)**, from Gross (2000), presents such a comparison, one that focused upon assessing the effectiveness of both atmospheric and oceanographic processes.

Inspection of this figure demonstrates, as shown earlier by Wilson and Haubrich (1976) and Wahr (1983), that the net effect of the sum of atmospheric wind and pressure fluctuations (determined on the basis of NCEP/National Center for Atmospheric Research (NCAR) reanalysis data) is insufficient to explain the Chandler excitation in terms of availability as compared to required power (whereas this excitation is entirely adequate insofar as the annual wobble is concerned as previously remarked). On the other hand, when the oceanographic excitation is added to the forcing at the Chandler period, with

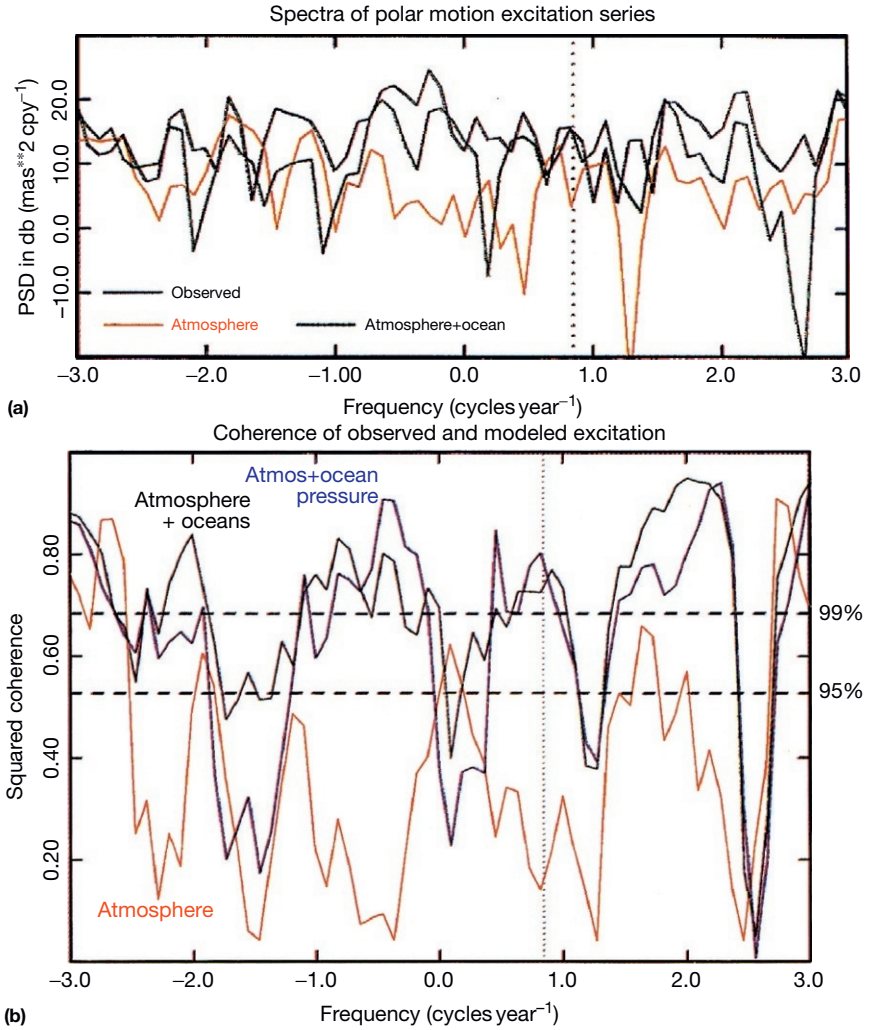


Figure 4 (a) The power spectral density (PSD) estimates in decibels (dB) computed from time series of polar motion excitation functions ($\chi(t)$) for the period 1985.0–1996.0 for the observed polar motion excitation function derived from space-geodetic Earth's rotation measurements (black curve), for the sum of the excitation functions due to atmospheric wind and pressure changes (red curve) where the atmospheric pressure term is that computed by assuming the validity of the inverted barometer approximation, and for the sum of all atmospheric and oceanic excitation processes (green curve). (b) Squared coherence between the observed polar motion excitation functions in the period 1985.0–1996.0 and the excitation functions due to the sum of atmospheric wind and (inverted barometer) pressure changes (red curve), due to the sum of atmospheric pressure and ocean bottom pressure fluctuations (blue curve), and due to the sum of all atmospheric and oceanic excitation processes. Reproduced from Gross RS (2000) The excitation of the Chandler wobble. *Geophysical Research Letters* 27: 2329–2342.

this determined on the basis of output of the MIT ocean general circulation model (Ponte et al, 1998) driven by 12 h wind stress and 24 h surface heat and freshwater fluxes from NCEP, the total power in the forcing at the Chandler period was found to fit the requirements of the observations extremely well.

In the Gross (2000) paper, this analysis in terms of the available power in the total forcing by the atmosphere and ocean was complemented by an analysis of the coherence between the observed excitation and modeled excitation, the results from which are reproduced here in Figure 4(b). Inspection of results of this analysis demonstrates that the atmospheric excitation alone is not coherent with the observed excitation, whereas the atmospheric plus oceanographic excitation is highly coherent with the observed excitation. Through detailed analysis, the author of this paper concludes that "the Chandler wobble is being excited by a combination of atmospheric and oceanic processes" and that "the single most important mechanism exciting the Chandler wobble has been ocean bottom pressure fluctuations which contribute about twice as much excitation power in the Chandler frequency band as do atmospheric pressure fluctuations." It would appear on the basis of these results that our understanding of the Chandler component of Earth's polar motion has become as well understood as are the similarly high-frequency variations in the l.o.d. discussed in the previous subsection.

In the following section of this chapter, I will extend the characteristic timescale of our interest in the Earth's rotation history to include the last million years, a period during which the surface of the solid Earth has been continuously subjected to a quasicyclic glaciation–deglaciation cycle involving extremely large exchanges of mass between the oceans and the surface of the continents on which massive continental ice sheets formed and episodically disintegrated. It should not be surprising, given the magnitude of this mass redistribution over the surface, that this 'ice-age cycle' has involved that it had, and continues to have, a profound impact upon the rotational state of the planet. What is often seen as very surprising, however, is the fidelity with which the Earth has conspired to remember the details of both the rotational excitation to which it was subjected and the response to it.

9.09.3 Theoretical Background II: The Viscoelastic Rotational Response to the Late Pleistocene Ice-Age Cycle

Given the extremely broad range of processes that contribute to the excitation of variations in the Earth's rotational state (see Figure 1), it is always an issue as to the degree to which one is able to uniquely ascribe the observed variability in a particular range of timescales to any one particular process. On the characteristic timescale of millennia and tens to hundreds of millennia, however, it is apparently the case that the excitation mechanism is so strongly dominated by the Late Pleistocene process of glaciation and deglaciation that such unambiguous attribution has come to be quite generally accepted. This is due to the advances that have been achieved over the past several decades in the development of a detailed viscoelastic theory of the GIA process.

9.09.3.1 The Global Theory of the GIA Process

The modern theory of global glacial isostasy addresses the question of the changes in the Earth's shape, gravitational field, and rotational state caused by the variations in surface mass load that occur when water is irreversibly removed from the oceans and delivered to the high-latitude continents to form large accumulations of land ice. Since the amount of water removed from the ocean basins in a typical 100 ky cycle of glaciation and deglaciation has been such as to cause mean sea level to fall by $\sim$120 m, it should not be surprising that this process has been accompanied by highly significant variations in the Earth's rotation, variations of sufficient amplitude as to leave indelible imprints in the geologic record of relative sea-level history. The dominant timescale of the Late Pleistocene ice-age cycle is most clearly appreciated on the basis of oxygen-isotopic measurements made on the calcium carbonate shells (tests) of foraminifera that are a ubiquitous component of the sedimentology of deep-sea sedimentary cores. One example of such a record is shown in Figure 5. The oxygen-isotopic anomaly $\delta^{18}O$, as measured on the tests of foraminifera extracted as a function of depth from such cores, provides a high-quality proxy for the amount of ice that existed on the continents as a function of time in the past, determined by the depth in the core from which the individual samples are taken (Shackleton, 1967). This is a simple consequence of the fact that the process whereby water is irreversibly removed from the oceans to build the ice sheets is a process that fractionates mass. $H_2{}^{16}O$ is preferentially removed from the oceans through the evaporation and precipitation processes whereby water is eventually delivered to the high-latitude continents as snow, resulting in an ocean left behind that is anomalously rich in the isotopologue $H_2{}^{18}O$ at a time in the past when large-scale continental ice sheets are in place, an anomaly that is recorded in the shells of foraminifera so long as these remain open systems. This isotopic anomaly is defined as

$$\delta^{18}O = \left(\frac{\left[{}^{18}O/{}^{16}O\right]_{\text{sample}}}{\left[{}^{18}O/{}^{16}O\right]_{\text{standard}}} - 1 \right) \qquad [9]$$

The record shown in Figure 5, from Ocean Drilling Program Site 677 in the Panama Basin off the west coast of the South American continent, remains one of the highest-resolution records available and was employed in Shackleton et al. (1990) to refine the entire chronology of the Pleistocene glacial epoch (see Section 9.09.7 in the succeeding text for a detailed discussion). Inspection of the power spectrum of this record, which is also shown in this figure for both the most recent and previous million years of the Pleistocene epoch, will show that for the most recent 900 000 years (see Deblonde and Peltier (1993) for a statistical assessment of the sharpness of this 'mid-Pleistocene climate transition'), the climate system has been dominated by a quasicyclic variation of ice cover with an average period of $\sim$100 000 years. Also evident by inspection of the power spectrum of this record are the additional weaker spectral peaks, one at a period of $\sim$41 000 years and a triplet of very weak spectral lines at periods of 19 000, 22 000, and 23 000 years. These additional spectral lines are associated with the component of ice-volume variability that is directly forced by the weak variations in solar radiation caused by the

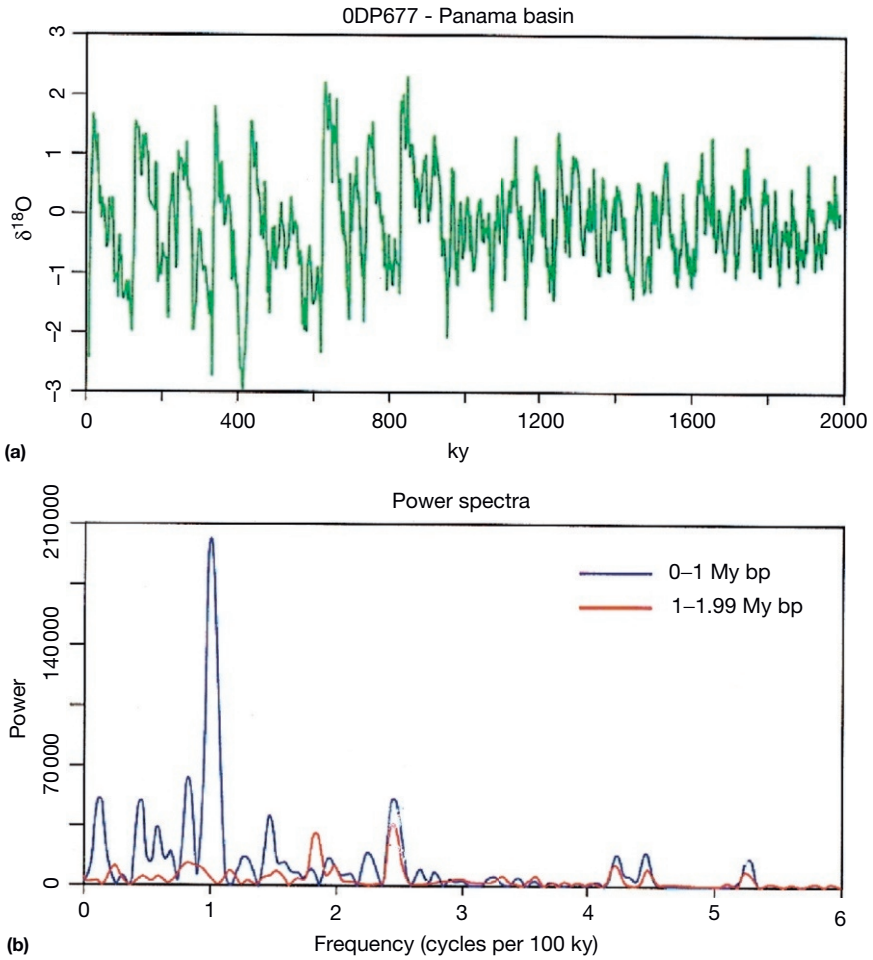


Figure 5 Oxygen-isotopic record from Ocean Drilling Program Site 677 in the Panama Basin extending to 2 million years before present from Shackleton et al. (1990). Also shown are the power spectra for the first and last million years of the record illustrating the onset of the 100 000-year cycle of glaciation and deglaciation in mid-Pleistocene time. Also evident in this continental ice-volume proxy record are the additional spectral lines associated with the variations of orbital obliquity and with the eccentricity modulation of the precessional effect (see **Section 9.09.7** for further discussion).

changing geometry of the Earth's orbit around the Sun that are in turn due to gravitational n-body effects in the solar system. Because the Earth and Sun are not alone in the solar system, the planet's orbit around the Sun does not consist of a time-invariant Keplerian ellipse but rather one whose ellipticity varies slowly with time at dominant periods of 100 000 and 400 000 years. Furthermore, as the planet continuously executes its almost elliptical orbit around the Sun, the axis, about which it spins and which presently makes an angle of $\sim 23.5^\circ$ with the plane of the ecliptic, also varies with time such that this angle (the so-called obliquity of the spin axis) varies with a dominant period of 41 000 years. As will be discussed more fully in **Section 9.09.7** of this chapter, it is these sources of variability in the Earth's rotational state together with the precession of the equinoxes that are responsible for the onset of the individual ice ages themselves.

Given the large variations in surface mass load that have accompanied the growth and decay of continental ice sheets over the past ~900 000 years of Earth history, during which a quasiperiodic ice-age cycle has persisted with a period of ~100 000 years, it should be clear that the key to understanding the excitation of the Earth's rotation that has accompanied the process must involve the ability to construct a detailed model of the variations in the moment of inertia tensor caused by the process. This requires the solution of what I have come to call the sea-level equation (SLE), an equation that accurately describes the way in which water must be distributed over the ocean basins when ice sheets melt so as to ensure that the evolving surface of the ocean remains a surface of constant gravitational potential. An initial version of this equation was first solved in papers by Clark et al. (1978) and Peltier et al. (1978) based upon the combination of results obtained in Peltier (1974, 1976), Farrell and Clark (1976), and Peltier and Andrews (1976). Since the completion of this original work, the theory underlying the SLE has been significantly extended, however, and it is these extensions that have proved crucial to the accurate understanding of the important role that the Earth's rotational response to the GIA process plays in the theory.

The solution of the SLE delivers a prediction of the history of relative sea-level change, $S(\theta, \lambda, t)$ say, that is caused by an assumed known history of the variations in continental ice

mass, represented by a history of continental ice-sheet thickness variations $I(\theta,\lambda,t)$. In these expressions, θ is latitude, λ is longitude, and 't' is time. The form of the SLE that relates these quantities is as follows:

$$S(\theta,\lambda,t)=C(\theta,\lambda,t)\left[\int_{-\infty}^{t}\mathrm{d}t'\iint_{\Omega}\mathrm{d}\Omega'\left\{L(\theta',\lambda',t')G_{\phi}^{L}(\phi,t-t')+\Psi^{R}(\theta',\lambda',t')_{R}G_{\phi}^{T}(\phi,t-t')\right\}+\frac{\Delta\Phi(t)}{g}\right] \quad [10]$$

In this equation, $C(\theta,\lambda,t)$ is the 'ocean function,' which is, as defined by Munk and Macdonald (1960), unity over the oceans and zero over the continents. The function 'C' is time-dependent because the removal of water from, or its addition to, the ocean basins causes the coastline to migrate, an effect that may be accurately computed on a global basis using the methodology described in Peltier (1994). The space–time-dependent surface mass load per unit area 'L' in the aforementioned equation has the composite form

$$L(\theta,\lambda,t)=\rho_{\mathrm{I}}I(\theta,\lambda,t)+\rho_{\mathrm{W}}S(\theta,\lambda,t) \quad [11]$$

in which ρ_{I} and ρ_{W} are the densities of ice and water, respectively. The angle ϕ in the Green functions G_{ϕ}^{L} and ${}_{R}G_{\phi}^{T}$ is the angular separation between the source point (θ',λ') and the field point (θ,λ). The function $\Psi^{R}(\theta,\lambda,t)$ is the variation of the centrifugal potential due to the changing rotational state of the Earth that is caused by the surface loading and unloading process associated with the ice-age cycle. Following Dahlen (1976), this may be expressed, to first order in perturbation theory, in terms of the following spherical harmonic expansion, as

$$\Psi^{R}(\theta,\lambda,t)=\Psi_{00}Y_{00}(\theta,\lambda,t)+\sum_{m=-1}^{+1}\Psi_{2m}Y_{2m}(\theta,\lambda,t) \quad [12]$$

with

$$\Psi_{00}=\frac{2}{3}\omega_{3}(t)\Omega_{\mathrm{o}}a^{2} \quad [13a]$$

$$\Psi_{20}=-\frac{1}{3}\omega_{3}(t)\Omega_{\mathrm{o}}a^{2}\sqrt{4/5} \quad [13b]$$

$$\Psi_{2,-1}=(\omega_{1}-i\omega_{2})\left(\Omega_{\mathrm{o}}a^{2}/2\right)\sqrt{2/15} \quad [13c]$$

$$\Psi_{2,+1}=-(\omega_{1}+i\omega_{2})\left(\Omega_{\mathrm{o}}a^{2}/2\right)\sqrt{2/15} \quad [13d]$$

The ω_i in the aforementioned equations are to be obtained as solutions to the appropriate version of eqns [3] in Section 9.09.2 in the preceding text, the form from which the externally applied torques τ_i have been removed as the rotational excitation in this case must be considered to be internally generated. The remaining terms in the SLE consist of the surface mass loading and tidal potential loading Green functions, which have been shown in Peltier (1976) and Peltier et al. (2012) to have the mathematical representations

$$G_{\phi}^{L}(\phi,t)=\frac{a}{m_{e}}\sum_{l=0}^{\infty}\left(1+k_{l}^{L}(t)-h_{l}^{L}(t)\right)P_{l}(\cos\theta) \quad [14a]$$

$${}_{R}G_{\phi}^{T}(\phi,t)=\frac{1}{g}\sum_{l=0}^{\infty}\frac{(2l+1)}{4\pi}\left(1+k_{l}^{T}(t)-h_{l}^{T}(t)\right)P_{l}(\cos\theta) \quad [14b]$$

in which k_l^T, h_l^T are the viscoelastic tidal potential loading Love numbers and k_l^L, h_l^L are the corresponding surface mass loading Love numbers of Peltier (1974). The renormalization factor $(2l+1)/4\pi$ in eqn [14b] is needed to ensure that the centrifugal potential introduces only local influence as discussed in detail in Peltier et al. (2012). Peltier (1976, 1985) had shown that the time-domain viscoelastic Love numbers that appear in eqns [14a] and [14b] may be expressed, in the case of impulsive point mass or centrifugal potential loading at the surface of the planet, in the form of the following normal-mode expansions:

$$k_{l}^{T}(t)=k_{l}^{T,E}\delta(t)+\sum_{j=1}^{M}q^{l}{}_{j}e^{-s_{j}^{l}t} \quad [15a]$$

$$h_{l}^{T}(t)=h_{l}^{T.E}\delta(t)+\sum_{j=1}^{M}r^{l}{}_{j}e^{-s_{j}^{l}t} \quad [15b]$$

$$k_{l}^{L}(t)=k_{l}^{L,E}\delta(t)+\sum_{j=1}^{M}q_{j}^{L}e^{-s_{j}t} \quad [15c]$$

$$h_{l}^{L}(t)=h_{l}^{L,E}\delta(t)+\sum_{j=1}^{M}r_{j}^{L}e^{-s_{j}^{l}t} \quad [15d]$$

In these normal-mode expansions, the $k_l^{T,E}, h_l^{T,E}, k_l^{L.E}$, and $h_l^{L,E}$ are the elastic surface mass load and tidal potential loading Love numbers of Farrell (1972), the s_j^l are the inverse relaxation times of a discrete set of normal modes of viscoelastic gravitational relaxation determined as the zeros of an appropriate secular function (Peltier, 1985) or by collocation (Peltier, 1974, 1976), and the amplitudes $q_j^l, r_j^l, q_j^l, r_j^l$ are the residues at these poles. Insofar as understanding of the polar wander component of the rotational response of the planet to the GIA process is concerned, the parameter k_2^T plays an especially crucial role as will be made clear in what follows. Solutions of the SLE are computed using an extension of the pseudo-spectral algorithm of Mitrovica and Peltier (1991).

9.09.3.2 Computation of the Rotational Response to Ice–Earth–Ocean Interactions in the Ice Age

The important role of the Love number $k_2^T(t)$ may be understood by returning to eqns [4] of Section 9.09.2 and focusing upon the polar wander component of the rotational response of the planet in circumstances in which it may be assumed that externally applied torques vanish ($\tau_i \equiv 0$) and noting that in this case, we must distinguish between two distinct but intimately related sources of excitation. The first of these derives from the perturbations of inertia caused by the direct affect of the isostatic adjustment process, an influence that may be represented in the form (e.g., Peltier, 1982)

$$I_{ij}^{\mathrm{GIA}}=\left(\delta(t)+k_{2}^{L}(t)\right)*I_{ij}^{R}(t) \quad [16]$$

in which the $I_{ij}^{R}(t)$ are the perturbations of inertia that would be obtained due to the variations in surface mass load if the Earth were a rigid body rather than viscoelastic and the symbol * represents the temporal convolution operation.

The second contribution to the perturbations of inertia is that due to the changing rotation itself, this may be derived from an application of a linearized version of MacCullagh's formula (e.g., see Munk and MacDonald (1960)) as

$$I_{13}^{\mathrm{ROT}} = \left(\frac{k_2^T * a^5 \omega_1 \omega_3}{3G}\right) = \left(\frac{k_2^T}{k_f}\right) * m_1 (C-A) \qquad [17a]$$

$$I_{23}^{\mathrm{ROT}} = \left(\frac{k_2^T * a^5 \omega_2 \omega_3}{3G}\right) = \left(\frac{k_2^T}{k_f}\right) * m_2 (C-A) \qquad [17b]$$

with

$$k_f = \left(\frac{3G}{a^5 \Omega_o^2}\right)(C-A) \qquad [17c]$$

in which $A = B$ has been assumed in eqns [4] and $k_2^T(t)$ is the tidal potential loading Love number of degree 2 defined in the preceding text, the parameter that will be seen to play a crucial role in what follows. Equally critical for the arguments to be presented is the so-called fluid Love number k_f, the value of which is determined entirely on the basis of the well-known equatorial flattening of the planet that is represented by the difference between the polar and equatorial moments of inertia $(C-A)$ in eqn [17c]. Substituting in eqn [17c] for Newton's gravitational constant 'G' the Earth's radius 'a' (for which we might take either the equatorial or polar value in order to investigate two hypothetical values for k_f), the present-day rate of angular rotation Ω_o, and the polar and equatorial moments of inertia C and A, respectively, taking all data from the tabulation of Yoder (1995) (see www.agu.org/references/geophys/4_Yoder.pdf), one obtains for the value of k_f, taking 'a' to be the polar radius, that

$$k_f \cong 0.9414 \qquad [18a]$$

or taking 'a' to be the equatorial radius, we obtain, alternatively,

$$k_f = 0.9382 \qquad [18b]$$

An important part of the discussion to follow will involve understanding of the connection between k_f and $\mathrm{Lim}_{t\to\infty} k_2^T(t)$, the infinite time asymptote of the viscoelastic tidal loading Love number of degree 2 in response to a constant tidal potential applied at $t = 0$ and maintained.

Since the solution of eqn [3c] for the change in the axial rate of rotation is uncomplicated, it will suffice to focus in what follows on the solution of eqns [3a] and [3b] for the polar wander component of the response to surface loading. This focus is furthermore motivated by the fact that the influence of polar wander strongly dominates the influence of rotational feedback upon sea-level history. Substituting eqns [17a] and [17b] together with eqn [16] into eqns [3a] and [3b] using [4a] and [4b] for the excitation functions, the Laplace-transform forms of the equations that follow are simply

$$s m_1 + \sigma\left(1 - \frac{k_2^T(s)}{k_f}\right) m_2 = \Psi_1(s) \qquad [19a]$$

$$s m_2 + \sigma\left(1 - \frac{k_2^T(s)}{k_f}\right) m_1 = \Psi_2(s) \qquad [19b]$$

where

$$\sigma = \Omega_o \frac{(C-A)}{A} \qquad [19c]$$

is the Chandler wobble frequency of the rigid Earth, 's' is the Laplace transform variable, and again, $A = B$ has been assumed. The Laplace-transformed forms of the excitation functions in eqns [19a] and [19b] are simply

$$\Psi_1(s) = \left(\frac{\Omega_o}{A}\right) I_{23}(s) - \left(\frac{s}{A}\right) I_{13}(s) \qquad [20a]$$

$$\Psi_2(s) = \left(\frac{\Omega_o}{A}\right) I_{13}(s) - \left(\frac{s}{A}\right) I_{23}(s) \qquad [20b]$$

with

$$I_{ij}(s) = \left(1 + k_2^L(s)\right) I_{ij}^{\mathrm{Rigid}}(s) \qquad [20c]$$

Now, eqns [19a] and [19b] are elementary algebraic equations for $m_1(s)$ and $m_2(s)$, and these may be solved, neglecting terms of order s^2/σ^2, as is appropriate for the analysis of a physical process evolving on a timescale that is much longer than the period of the Chandler wobble, to obtain

$$m_1(s) = \left(\frac{\Omega_o}{A\sigma}\right)\left[\frac{1 + k_2^L(s)}{1 - \dfrac{k_2^T(s)}{k_f}}\right] I_{13}^{\mathrm{Rigid}}(s) = H(s) I_{13}^{\mathrm{Rigid}}(s) \qquad [21a]$$

$$m_2(s) = H(s) I_{23}^{\mathrm{Rigid}}(s) \qquad [21b]$$

A convenient shorthand form for the solution vector $(m_1, m_2) = \underline{m}$ is to write

$$\underline{m}(s) = \frac{\underline{\Psi}^L(s)}{\left[1 - \dfrac{k_2^T(s)}{k_f}\right]} = H(s)\left(I_{13}^{\mathrm{Rigid}}(s), I_{23}^{\mathrm{Rigid}}(s)\right) \qquad [22a]$$

where

$$\underline{\Psi}^L(s) = \left[\left(\frac{\Omega_o}{A\sigma}\right)\left(1 + k_2^L(s)\left(I_{13}^{\mathrm{Rigid}}(s), I_{23}^{\mathrm{Rigid}}(s)\right)\right)\right] \qquad [22b]$$

From eqns [22], it will be clear that the polar wander solution $\underline{m}(s)$ will depend critically upon the ratio $k_2^T(s)/k_f$. This fact was more fully exposed in the analysis of Peltier (1982) and Wu and Peltier (1984) who rewrote the Laplace-transformed domain forms of $k_2^T(s)$ and $k_2^L(s)$ as (e.g., see eqn [61] of Wu and Peltier (1984))

$$k_2^T(s) = k_2^T(s=0) - s\sum_{j=1}^{N} \frac{\left(q_j'/s_j\right)}{(s+s_j)} \qquad [23a]$$

$$k_2^L(s) = (-1 + \ell_s) - s\sum_{j=1}^{N} \frac{(q_j/s_j)}{(s+s_j)} \qquad [23b]$$

in which the superscript $\ell = 2$ on q_j^2, r_j^2, s_j^2 has been suppressed for convenience. Substituting eqn [23a] into eqn [22], this may be rewritten as

$$\underline{m}(s) = \frac{\underline{\Psi}^L(s)}{\left[1 - \dfrac{k_2^T(s=0)}{k_f}\right] + \dfrac{s}{k_f}\displaystyle\sum_{j=1}^{N} \frac{\left(q_j'/s_j\right)}{(s+s_j)}} \qquad [24]$$

In discussing the formal inversion of eqn [24] into the time domain, I will distinguish for the purpose of this chapter the

results that follow from the application of an important modified form of this equation from those that follow from the complete eqn [24].

9.09.3.3 The 'Equivalent Earth Model' Approach of Munk and MacDonald (1960): An Approximate Solution

Since the surface of the Earth is broken into a large number of individually rigid lithospheric 'plates' whose boundaries are in general weak, it should be the case that, at the large horizontal scale represented by spherical harmonic degree 2, the effective $k_2^T(s=0)$ will be close to k_f since in this limit, the absence of strength at plate boundaries will enable the planet as a whole to adjust to the tidal (rotational) forcing as if the planet had no surface lithosphere at all. This is highly plausible on physical grounds as the product $\mathrm{Lim}_{s\to 0} s k_2^T(s)$, via the Tauberian theorem (e.g., see Widder (1946)), is identical to $\mathrm{Lim}_{t\to\infty} k_2^T(t)$. That $k_2^T(s=0)$ does tend towards k_f in response to the sustained application of the rotational forcing as the thickness of the lithosphere tends to zero will be clear on the basis of Figure 6 (see also Table 5 in Wu and Peltier (1984)). Based upon Figure 6, it will be clear that the deviation of $k_2^T(s=0)$ from k_f is by $<1\%$ for either of the plausible values of k_f denoted by the dashed lines. It is an interesting question as to whether this small deviation is physical as opposed to being an artifact due to a slight error in the radial variation of density in the PREM model that is being employed to represent the unperturbed form of the spherically symmetrical approximation to the planet upon which the GIA process is superimposed. For present purposes, I will consider that the difference is plausibly physically meaningful and derivative of the action of the mantle convection process. The mathematical methods required to invert eqn [24] under the assumption

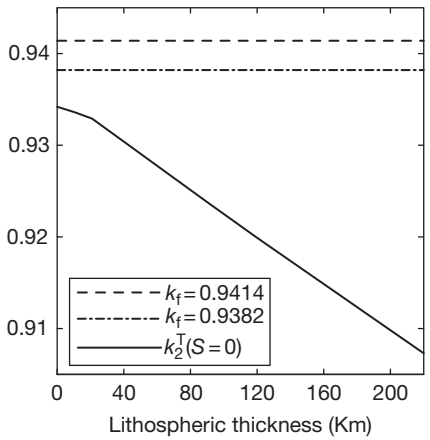


Figure 6 The infinite time asymptotic value of the tidal potential loading Love number of degree 2 is shown as a function of lithospheric thickness. This infinite time limit is identical to the limit in which the Laplace transform variable 's' equals zero. Also shown in the figure are two plausible values of the fluid Love number k_f, which is a nondimensional constant determined by the observed flattening of Earth's shape as measured by the difference between the polar and equatorial moments of inertia (C − A). Note that the difference between $k_2^T(s=0)$ and k_f in the limit of zero lithospheric thickness is $<1\%$ for either one of the two values of k_f. See text for further discussion.

$$1 - \frac{k_2^T(s=0)}{k_f} \equiv 0 \qquad [25]$$

were presented in Peltier (1982) and Wu and Peltier (1984) and will not be repeated here. This is the assumption upon which the equivalent Earth model (EQM) approach of Munk and MacDonald (1960) is based. However, it will be important to make the point immediately that eqn [24] subject to eqn [25] does not constitute a mathematically unstable formulation as it remains finite in the infinite time limit by the Tauberian theorem mentioned in the preceding text, contrary to the apparent claim made by Mitrovica et al. (2005).

9.09.3.4 A Formulation that Retains the Difference Between $k_2^T(s=0)$ and k_f

In this case, the assumption $k_2^T(s=0) = k_f$ that is at the heart of the 'equivalent Earth model' approach is abandoned, and the Laplace transform domain impulse response is then

$$H(s) = \left(\frac{\Omega_o}{A\sigma}\right) \frac{1 + k_2^L(s)}{\frac{s}{k_f}\sum_{j=1}^{N} \frac{(q_j/s_j)}{(s+s_j)} + \varepsilon} \qquad [26a]$$

where

$$\varepsilon = 1 - \frac{k_2^T(s=0)}{k_f} \qquad [26b]$$

As will become clear, even though ε is a small quantity (especially in the case that the finite thickness of the lithosphere may be neglected in the limit $t\to\infty$), retaining it in expression [26a] for the impulse response could have a significant impact upon the solution as the rotational response of the system is modified. Now, the construction of the solution for the time-domain form of the impulse response $H(t)$ proceeds in this case as in the case based upon the EQM assumption, although the result differs from a physical perspective, especially in circumstances in which, as we will see, a perturbation to the inertia tensor is applied and maintained over a sufficiently long timescale. In this case, it is useful to make the distinction between the Chandler wobble frequency of a rigid Earth σ and the Chandler wobble frequency of the viscoelastic Earth σ_o, by employing the definition

$$\sigma_o = \frac{(k_2^T(s=0) - k_2^{TE})}{k_2^T(s=0)}\sigma \qquad [27]$$

We may then rewrite the expression for $H(s)$ as

$$H(s) = \left(\frac{\Omega_o}{A\sigma_o}\right) \frac{(1 + k_2^L(s))}{\left((1-\varepsilon)s\sum_{i=1}^{N} \frac{g_j}{s+s_j}\right) + \varepsilon'} \qquad [28a]$$

with

$$\varepsilon' = \varepsilon\frac{\sigma}{\sigma_o} \qquad [28b]$$

and,

$$g_j = \frac{q_j'/s_j}{\sum_j (q_j'/s_j)} \qquad [28c]$$

The inversion of $H(s)$ into the time domain now proceeds by expanding the sum in the denominator of eqn [28a] in the form

$$\sum_{j=}^{N}\frac{g_j}{(s+s_j)}=\frac{Q_{N-1}(s)}{\prod_{j=1}^{N}(s+s_j)}=\frac{\prod_{j=1}^{N-1}(s+\lambda_j)}{\prod_{j=}^{N}(s+s_j)} \qquad [29]$$

since $\sum g_j \equiv 1$. Then, we have, suppressing for the moment the factor $(\Omega_o/A\sigma_o)$,

$$H(s)=\frac{\prod_{j=1}^{N}(s+s_j)\left[1+k_2^L(s)\right]}{(1-\varepsilon)s\prod_{i=1}^{N-1}(s+\lambda_i)+\varepsilon'\prod_{j=1}^{N}(s+s_j)} \qquad [30]$$

Now, substituting for the function $1+k_2^L(s)$ from eqn [23b], we obtain

$$H(s)=\frac{\prod_{j=1}^{N}(s+s_j)\ell_s}{(1-\varepsilon)s\prod_{i=1}^{N-1}(s+\lambda_i)+\varepsilon'\prod_{i=1}^{N}(s+s_i)}$$

$$+\sum_{j=1}^{N}\frac{(-q_j/s_j)s\prod_{i\neq j}^{N}(s+s_i)}{(1-\varepsilon)s\prod_{i=1}^{N-1}(s+\lambda_i)+\varepsilon'\prod_{i=1}^{N}(s+s_i)} \qquad [31a]$$

$$=\frac{\prod_{j=1}^{N}(s+s_j)\ell_s}{(1-\varepsilon+\varepsilon')\prod_{i=1}^{N}(s+\kappa_i)}+\sum_{j=1}^{N}\frac{(-q_j/s_j)s\prod_{i\neq j}^{N}(s+s_j)}{(1-\varepsilon+\varepsilon')\prod_{i=1}^{N}(s+\kappa_i)} \qquad [31b]$$

where the κ_i are the N roots of the polynomial in the denominator of the two terms in eqn [31a]. This expression for the impulse response may be further reduced by rewriting the ratios of products as follows:

$$\frac{\prod_{j=1}^{N}(s+s_j)}{\prod_{j=1}^{N}(s+\kappa_i)}=1-\frac{q'(s)}{\prod_{i=1}^{N}(s+\kappa_i)} \qquad [32a]$$

where now,

$$q'(s)=\prod_{j=1}^{N}(s+\kappa_i)-\prod_{j=1}^{N}(s+s_j) \qquad [32b]$$

and

$$\frac{s\prod_{i\neq j}^{N}(s+s_i)}{\prod_{j=1}^{N}(s+\kappa_i)}=1-\frac{R_j'(s)}{\prod_{i=1}^{N}(s+\kappa_i)} \qquad [33a]$$

with

$$R_j'(s)=\prod_{i=1}^{N}(s+\kappa_i)-s\prod_{i\neq j}^{N}(s+s_i) \qquad [33b]$$

We then have, for the Laplace transform of the impulse response, the expression

$$H(s)=\frac{\ell_s}{(1-\varepsilon+\varepsilon')}\left\{1-\frac{q'(s)}{\prod_{i=1}^{N}(s+\kappa_i)}\right\}$$

$$+\frac{1}{(1-\varepsilon+\varepsilon')}\sum_{j=1}^{N}\left(-\frac{r_j}{s_j}\right)\left\{1-\frac{R_j'(s)}{\prod_{i=1}^{N}(s+\kappa_i)}\right\} \qquad [34a]$$

or

$$H(s)=\frac{\ell_s-\sum_{j=1}^{N}r_j/s_j}{(1-\varepsilon+\varepsilon')}-\frac{\ell_s q'(s)}{(1-\varepsilon+\varepsilon')\prod_{i=1}^{N}(s+\kappa_i)}$$

$$+\frac{1}{(1-\varepsilon+\varepsilon')}\sum_{j=1}^{N}\frac{(q_j/s_j)R_j'(s)}{\prod_{i=1}^{N}(s+\kappa_i)} \qquad [34b]$$

Denoting $\ell_s-\sum_{j=1}^{N}r_j/s_j=1+k_2^{LE}=D_1$, say, then we may further reduce the expression for the impulse response to

$$H(s)=\frac{D_1}{(1-\varepsilon+\varepsilon')}-\frac{1}{(1-\varepsilon+\varepsilon')}\left\{\frac{\ell_s q'(s)-\sum_{j=1}^{N}(q_j/s_j)R_j'(s)}{\prod_{i=1}^{N}(s+\kappa_i)}\right\} \qquad [35]$$

The inverse Laplace transform of this expression is such that the solution in the present case, in which $k_2^T(s=0)\neq k_f$, is just

$$m_1(t)=\frac{1}{(1-\varepsilon+\varepsilon')}\left(\frac{\Omega_o}{A\sigma_o}\right)$$

$$\left\{\left[\ell_s-\sum_{j=1}^{N}\frac{r_j}{s_j}\right]I_{13}^{\text{Rigid}}(t)+\sum_{i=1}^{N}E_i'e^{-\kappa_i t}*I_{13}^{\text{Rigid}}(t)\right\} \qquad [36a]$$

$$m_2(t)=\frac{1}{(1-\varepsilon+\varepsilon')}\left(\frac{\Omega_o}{A\sigma_o}\right)$$

$$\left\{\left[\ell_s-\sum_{j=1}^{N}\frac{r_j}{s_j}\right]I_{23}^{\text{Rigid}}(t)+\sum_{i=1}^{N}E_i'e^{-\kappa_i t}*I_{23}^{\text{Rigid}}(t)\right\} \qquad [36b]$$

where

$$E_i'=\left\{-\ell_s q'(-\kappa_i)+\sum_{j=1}^{N}\frac{r_j}{s_j}R_j'(-\kappa_i)\right\}/\prod_{i=j}^{N}(\kappa_j-\kappa_i) \qquad [36c]$$

The polar wander velocity vector components are obtained simply by time differentiation of eqns [36a] and [36b]. It is useful to compare the results in eqn [36] to those obtained under the EQM assumption. In the limit $\varepsilon \to o$, we have $\kappa_N=0$ and $\kappa_i=\lambda_i$, which are the $N-1$ relaxation times that govern the system in this limit (see Wu and Peltier (1984)). In this case, the parameter E_N' in the in the preceding text becomes

$$E_N'=-\frac{\ell_1 q(o)}{\prod_{j=1}^{N-1}(\kappa_j-\kappa_N)}=-\frac{\ell_s q(o)}{\prod_{j=1}^{N-1}\lambda_i} \qquad [37]$$

and the EQM result is fully recovered. The mathematical analysis leading to the aforementioned analytic result for the polar

wander speed prediction embodied in eqn [36] is original to Peltier and Luthcke (2009) from which paper the aforementioned analysis has been extracted.

In order to compare the temporal histories of the rotational anomalies predicted in the two cases in response to an ice-age cycle of loading and unloading, it will be important to proceed by keeping as many features of the Earth model fixed as possible. To this end and for much of the rest of this chapter, I will focus primarily upon the VM2 viscosity model of Peltier (1996b) and the simple multilayer approximation to it that has been referred to as VM5a by Peltier and Drummond (2010) that will be employed together with the PREM radial elastic structure of Dziewonski and Anderson (1981). However, I will also discuss the results for simple two-layer viscosity models that may be directly compared to the earlier results for such models (e.g., see Peltier and Jiang (1996)).Equally important, of course, will be the model of glaciation and deglaciation that is employed to represent the surface mass load forcing. This is described in the next subsection.

9.09.3.5 Models of the History of Variations in the Elements of Earth's Moment of Inertia Tensor

As stated in the preceding text, the surface mass load forcing associated with the ice-age cycle of the Late Quaternary will, for the most part, be taken to be that either of the ICE-5G (VM2) model of Peltier (2004) or of the ICE-6G (VM5a) model of Peltier et al. (2014) that has been designed to eliminate the flaws in the North American and Antarctic components of the loading history identified in Argus and Peltier (2010) and Argus et al. (2011). **Figure 7** shows the distribution of the continental ice sheets at the Last Glacial Maximum (LGM) 26 000 years ago compared with that in the modern climate system of the Holocene along with the detailed variations of the vertical mean buoyancy frequency of the oceans for these two epochs according to the detailed reconstructions of atmosphere–ocean climate for these two epochs described in Peltier and Solheim (2004). **Figure 8** compares the contribution to the eustatic rise of sea level from LGM to the present for the ICE-5G (VM2) model with that of the ICE-6G-A (VM5a) refined model for each of the primary regions from which significant deglaciation occurred subsequent to LGM. ICE-6G is characterized by a significant geographic redistribution of the surface mass load. The most important of these changes, which the interested reader will find discussed in detail in Peltier et al. (2014); Argus et al. (2014), involves a significant redistribution of mass over the North American continent and a diminution of the mass loss from Antarctica during post-LGM deglaciation and the elimination of significant mass loss from the interior of East Greenland. In both models, meltwater pulse 1A in the Barbados record (see the succeeding text) is assumed to derive primarily from the northern hemisphere although ICE-6G envisions a small contribution to this pulse from Antarctica as well. Meltwater pulse 1B in the Barbados record is assumed to derive entirely from Antarctica in both models. It should also be noted that the ~128 m net

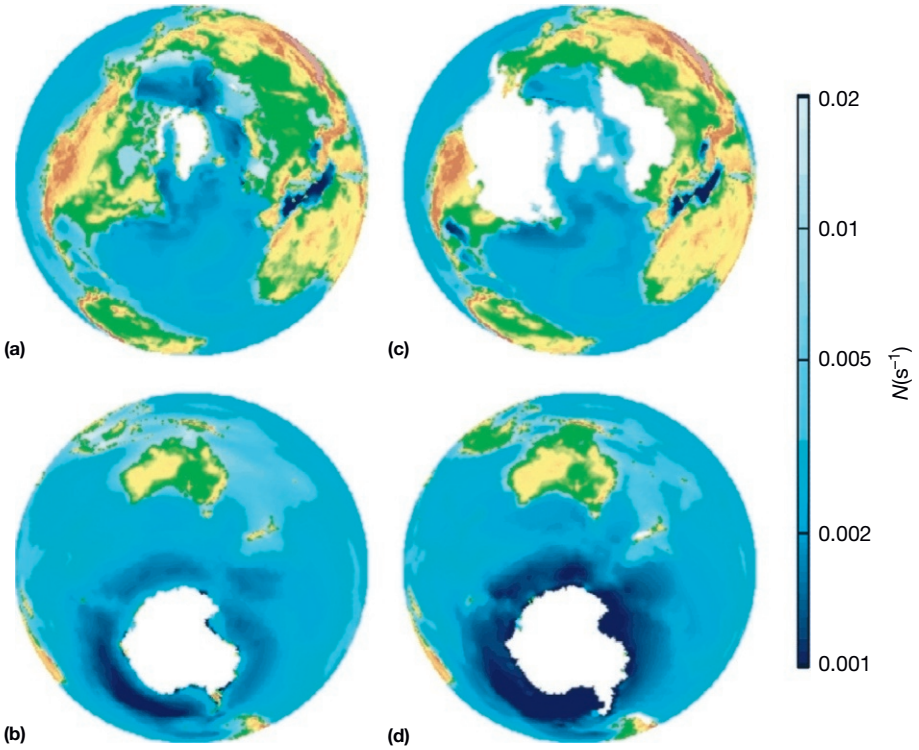


Figure 7 The difference between the distribution of land ice at LGM and the vertically averaged static stability (Brunt–Väisälä frequency '*N*') of the LGM ocean as compared with the modern ocean. The latter data are provided by the coupled atmosphere–ocean model integrations for the Last Glacial Maximum period of Peltier and Solheim (2004) and for the modern ocean by the same model operating under modern boundary conditions. These data are required for the reconstruction of the tidal regime by direct numerical solution of Laplace's tidal equation. Reproduced from Griffiths SD and Peltier WR (2008) Mega-tides in the glacial ocean and rapid climate change. *Geophysical Research Letters* 35: L08605, http://dx.doi.org/10.10292008GL033263.

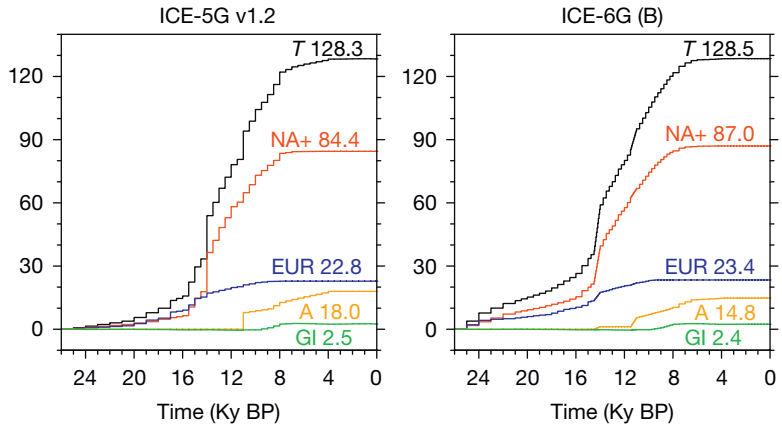


Figure 8 Ice-equivalent eustatic sea-level curves for the ICE-5G (VM2) and the ICE-6G (VM5a) models of the last glacial–interglacial transition and the contributions from each of the main regions from which land ice was lost. By ice-equivalent eustatic sea level, I mean the rise of sea level expected on the basis of the volume of water produced by the melting ice divided by the present-day surface area of the oceans. If the variation in the surface area of the oceans were taken into account, the glacial–interglacial eustatic rise would increase by ~10 m (e.g., see Peltier (2007a) for a detailed discussion). For the purpose of constructing this figure, I have left out the contributions from the Patagonian ice sheet and from the mountain glaciers of New Zealand, which explains why the sum over all regional ice masses does not exactly equal the total. This is done so as to provide a specific number for the contribution to eustatic sea-level rise for Antarctica with adding needless clutter to the diagrams. As discussed fully in the text, the sharp onset of Antarctic deglaciation at meltwater pulse 1B time ~11 500 years ago is constrained by data from sedimentary cores raised from the Antarctic shelf (see Figure 10 in the succeeding text and the discussion of these data in MacKintosh et al. (2011)). Notable is that in both of these reconstructions, LGM has been assumed to have occurred at ~26 000 years before present rather than the conventionally assumed age of 21 000 years ago. The data on the basis of which this revision is required were presented in Peltier and Fairbanks (2006). Of further importance is that the net rise of ice-equivalent eustatic sea level exceeds the conventionally assumed 120 m only because the timing of LGM has been pushed back to 26 000 years before present.

eustatic rise during deglaciation is greater than the usually assumed maximum of ~120 m only because the timing of LGM has been assumed to have been reached by 26 000 years before present rather than by the usually assumed time of 21 000 years before present. In both models, the sea-level depression at this conventional time is actually somewhat less than 120 m.

Since the relative sea-level data set from the island of Barbados has been employed to validate the total surface mass load, it will be useful to consider the extent to which the ICE-6G (VM5a) model of Peltier et al. (2014) reconciles the observed sea-level history at this site. **Figure 9**, which presents the extended compilation of Barbados data from Peltier and Fairbanks (2006), compares this observed RSL history with that predicted using the refined ICE-6G (VM5a) model of the GIA process and the SLE formalism as the red curve. On the same figure and shown as the red step discontinuous curve, the ice-equivalent eustatic sea-level history for this model is shown. The LGM ice-equivalent eustatic low stand of the sea in the ICE-6G (VM5a) model, assuming the modified LGM age of 26 ka proposed by Peltier and Fairbanks (2006) for the time of deepest glaciation, is ~128 m. This exceeds the usually assumed depth of the eustatic low stand of the sea, as previously discussed, only because the timing of LGM has been pushed back to 26 ka from the usually assumed age of 21 ka. The necessity of this reinterpretation has been defended in Peltier and Fairbanks (2006). By the ice-equivalent eustatic curve, I mean the variation of globally averaged sea level deduced directly on the basis of the time-dependent ratio of the volume of water generated by melting ice, taking into account the difference in density between ice and water, divided by the *modern* surface area of the oceans. This purposefully neglects the impact of the changing surface area of the oceans that accompanies the ice-age cycle itself so that the change in eustatic level is solely a function of the changing ice volume. The value of 128 m is still extremely close to the conventional oxygen isotope-derived estimate of ~120 m (e.g., see Shackleton (2000)), which was based upon the assumption that the depth of the LGM low stand was being accurately measured by the Barbados data itself. In the recent paper of Oba and Irino (2012), further evidence is provided of the validity of this depth of the eustatic low stand of the sea. It is a further important property of the extended Barbados record of Peltier and Fairbanks that the data appear to strongly reject the hypothesis of Yokoyama et al. (2000) that a strong meltwater pulse occurred at 19 ka. In the ICE-6G reconstruction, such rapid melting events are taken to have occurred only at the times of meltwater pulses 1A and 1B originally identified by Fairbanks (1989) of respective ages of 14.2 and 11.5 ka and which are clearly evident in **Figure 9**.

Also evident by inspection of the previous **Figure 8** is that the contribution to ice-equivalent eustatic sea-level rise from the Antarctic continent in version 1.2 of the ICE-5G (VM2) and ICE-6G (VM5a) models is characterized by a significant abrupt rise at ~11 500 years before present. This late onset of Antarctic melt-back has always been a characteristic of the ICE-NG sequence of models, it having been argued (e.g., see Peltier (2005)) that this region was the most likely source of the meltwater pulse 1B event that is clearly evident in the extended Barbados sea-level record discussed in Peltier and Fairbanks (2006). Very recently, this assumption in the ICE-NG reconstructions has been strikingly verified by the dating of an hiatus that has been observed to characterize the sedimentary

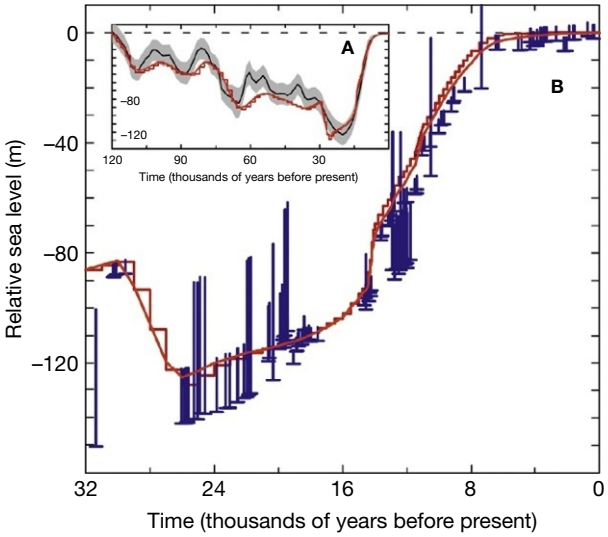


Figure 9 The fit of the predicted relative sea-level history at the island of Barbados to the extended coral-based data set from this location tabulated in Peltier and Fairbanks (2006) for model ICE-6G (VM5a). The blue symbols with error bars of various lengths represent these new Barbados data. Those data represented by the shorter error bars of 5 m length derive from the *Acropora palmata* species of coral that provide the best constraints on sea level. The data represented by the error bars on intermediate 20 m length derive from the *Montastrea annularis* species of coral. The data represented by the longest error bars derive from either the *Porites astreoides* species or the *Diploria* species. The inset to the figure shows the comparison between the eustatic history of the ICE-6G model and the complete 10^5-year glacial cycle with that inferred by Waelbroeck et al. (2002) based upon benthic δ^{18}O records corrected for the influence of the change in abyssal ocean temperature. It is notable that the GIA-based and paleoceanographic models are very close in terms of eustatic sea level.

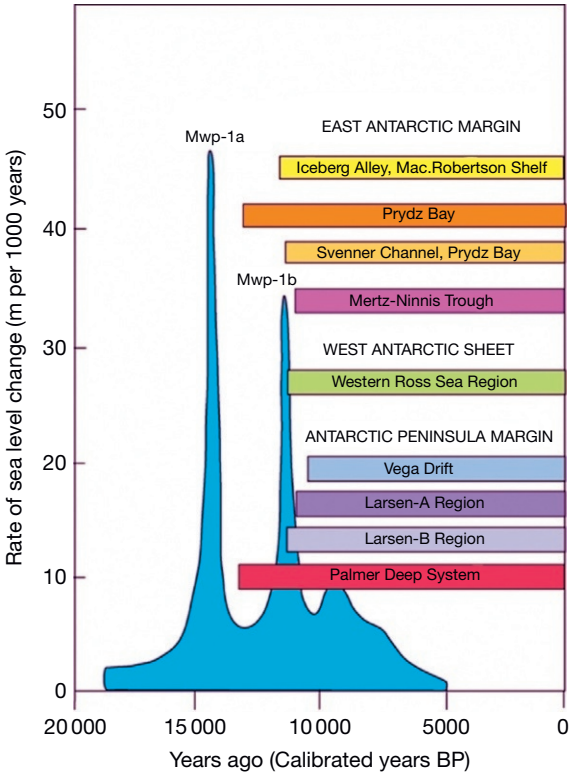


Figure 10 The timing of the onset of pullback of Antarctic ice from the shelf break surrounding the continent determined on the basis of the age of the hiatus in each of the sedimentary cores that marks the transition to the onset of continuous marine sedimentation following a period during which such sedimentation was absent. The data are those from the paper of MacKintosh et al. (2011) and the figure is redrawn from one provided to this author by Eugene Domack. Notable is the fact that at almost all locations from which such sedimentary cores have been raised, the timing of the onset of ice loss from Antarctica is coincident with meltwater pulse 1B in the Barbados record. These data directly refute the notion that any significant contribution to the earlier meltwater pulse 1A event could have originated from Antarctica.

stratigraphy of a very large number of sedimentary cores that have been raised from the shelf outboard of the current coastline of Antarctica. The sedimentology above the hiatus in each core is characterized by continuous marine sedimentation, whereas below the hiatus, marine sedimentation is absent. At each of the sites of these sedimentary cores, the hiatus is interpreted as representing the time at which the grounded ice, which extended out to the shelf break at LGM, first pulled back towards the present coastline (e.g., See Domack et al. (2005) and MacKintosh et al. (2011)). The age of the hiatus is therefore the age of the initial onset of significant Antarctic deglaciation. Inspection of the available data (from the paper by MacKintosh et al. (2013)) shown in **Figure 10** demonstrates that the time of onset of deglaciation is effectively coincident with the timing of meltwater pulse 1B in the Barbados record of sea-level history. These new observational results effectively rule out the suggestion in Clark et al. (2002) that the primary source of the earlier and more intense meltwater pulse 1A event was Antarctica. Peltier (2005) discussed additional evidence on the basis of which this suggestion appears to be implausible, and Tarasov and Peltier (2005, 2006) had provided further discussion of the Clark et al. (2002) suggestion. A sequence of predictions of relative sea-level histories based upon the ICE-5G (VM2) model at a sequence of well-dated sites from Scotland derivative of isolation basin-based inferences of RSL (see Peltier et al. (2002) for a detailed discussion of these data) appears to further reinforce the hypothesis of an Antarctic origin for meltwater pulse 1B. As discussed in Peltier et al. (2002), the highly nonmonotonic nature of these RSL curves, which are so well fit by the theoretical predictions (see also **Figure 23** in the succeeding text for a single example of these records from the Arisaig site), is significantly a consequence of the assumed late glacial melting event emanating from Antarctica and its abrupt onset as is characteristic of the ICE-5G v1.2 and ICE-6G (VM5a) model reconstructions.

The primary models of the glaciation–deglaciation process that will be employed here for the purpose of analyzing the Earth's rotational response to the ice-age cycle are therefore ICE-5G (VM2) and the closely related ICE-6G (VM5a) models, the extension of which in the interval between the Eemian interglacial and LGM is based upon the SPECMAP record of Imbrie et al. (1984). In the inset to **Figure 9**, the complete ICE-6G record of the most recent glaciation–deglaciation cycle is compared with the reconstruction of Waelbroeck et al. (2002), whose reconstruction was based upon deep-sea δ^{18}O records

corrected for the influence of the temperature variability of the abyssal ocean. The comparison of this record with that for the ICE-6G (VM5a) model demonstrates that the two approximations to the ice-equivalent eustatic sea-level history over the most recent glaciation–deglaciation cycle are extremely similar. In order to construct a complete surface loading model for use in the computation of the Earth's rotation anomalies, however, a model that includes more than a single glacial cycle is required. Since approximately seven such cycles have occurred during the past 800 000 years of Earth history, for the purpose of the analyses to be discussed herein, it will be assumed that an ~100 ky periodic cycle of glaciation and deglaciation has continued to operate over this same period.

Construction of the polar wander and length-of-day solutions requires the time series $I_{13}^{\text{Rigid}}(t)$, $I_{23}^{\text{Rigid}}(t)$, and $I_{33}^{\text{Rigid}}(t)$, inputs to the calculation that may be computed from the definition of the I_{ij}^{Rigid}. This definition is

$$I_{ij}^{\text{Rigid}}(t) = \iint \nu(\theta, \phi, t)\left(a^2\delta_{ij} - x_i x_j\right) \mathrm{d}s \qquad [38]$$

where the integral is over the surface of the sphere and where $\nu(\theta, \phi, t)$ is the surface mass load per unit area. Since this may be expressed in the form of a spherical harmonic expansion as

$$\nu(\theta, \phi, t) = \sum_{\ell=o}^{\infty} \sum_{m=-\ell}^{+\ell} \nu_{\ell m}(t) Y_{\ell m}(\theta, \phi) \qquad [39]$$

it follows that

$$I_{13}^{\text{Rigid}}(t) + iI_{23}^{\text{Rigid}}(t) = -\left(\frac{32}{15}\right)^{1/2} \pi a^4 \nu_{21}(t) \qquad [40a]$$

$$I_{33}^{\text{Rigid}}(t) = -\left(\frac{8}{3}\right)\pi a^4 \left[\left(\frac{1}{5}\right)^{1/2} \nu_{20}(t) - 2\nu_{oo}(t)\right] \qquad [40b]$$

An issue that immediately arises concerning the evaluation of these inputs to the theory concerns the way in which the component of the surface load is computed that is associated with the water that is extracted from or added to the oceans as the ice sheets grow and decay, respectively. Although the loading and unloading of the continents due to the growth and decay of the ice sheets are an a priori-specified input to the theory, the ocean loading component must be obtained self-consistently as a solution to the SLE eqn [10]. However, this equation for the space–time history of the load on the ocean basins includes the influence of rotational feedback through the convolution of Ψ^R with ${}_R G_\phi^T$. The problem of determining the rotational response to the ice-age cycle must therefore be determined iteratively. We first compute a solution for the ω_i by neglecting this feedback effect. We next add the feedback terms into eqn [10] and recompute the sea-level history and thus the modification to the ocean load needed to determine the inertia perturbations required for the computation of the $\omega_i(t)$, etc. The iteration sequence is found to converge rapidly. It is also important to note that the ocean component of the surface load varies over a single glacial cycle in a way that depends somewhat upon the viscosity model of the Earth that is employed to represent the GIA process.

Figure 11 illustrates two example solutions for the time series of the Cartesian components of the angular velocity vector for two different combinations of ice-sheet loading history and mantle viscosity profile. In Figure 11(a), the time series are shown for models in which seven 100 ky cycles of glacial loading and unloading are assumed. The time series plotted are of $m_1(t) = \omega_1(t)/\Omega_o$ and $m_2(t) = \omega_2(t)/\Omega_o$ in which Ω_o, as before, is the angular velocity of Earth about its unperturbed axis of rotation. Notable is the fact that there exists a secular variation in each component upon which the impact of individual 100 000-year periodic ice-age cycles is superimposed. The model of glaciation and deglaciation for ICE-5G (VM2) is exactly that illustrated in Figure 5 of Peltier and Luthcke (2009). The second model is denoted ICE-6G (VM5aD1), the viscosity model of which, VM5aD1, is one of the soft D″ (VM5a) models described in Peltier and Drummond (2010), whereas the ice loading and unloading history is a slightly modified version of ICE-5G in which the only significant changes have involved the slight redistribution of mass

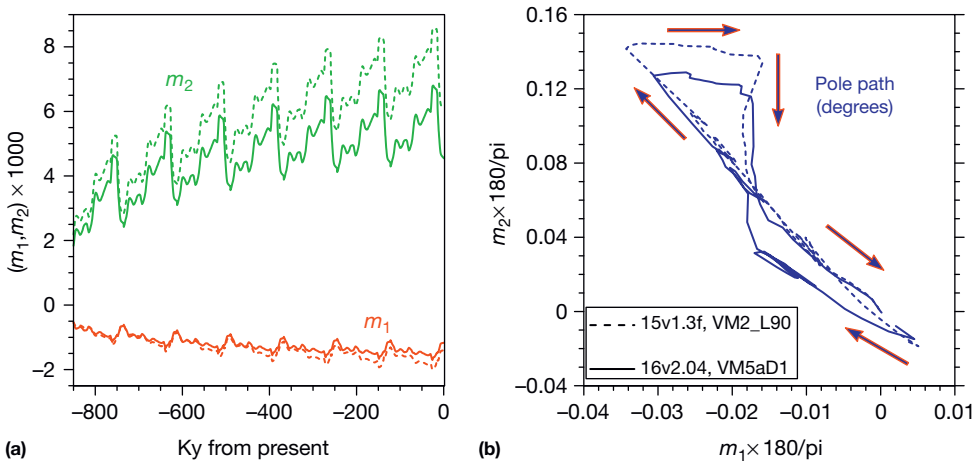


Figure 11 The reaction of the rotation pole in response to forcing by the 100 ky cycle of Late Pleistocene time. (a) The response of the pole in terms of the variables m_1 and m_2 for two different models of the inertia tensor variations for an assumed seven cycles of glacial loading and unloading. The two Earth models for which results are shown are ICE-5G (VM2) and ICE-6G (VM5a). (b) The actual pole paths for these two models are also shown in which the 'loops' associated with the two meltwater pulses are clearly evident.

over North America required to eliminate the misfits to the large database of GPS observations of vertical motion discussed in Argus and Peltier (2010) and a modification to the Antarctic component of the deglaciation history, which essentially eliminates mass loss from the interior of East Antarctica and moderately redistributes mass loss from West Antarctica (see Argus et al. (2014) and Peltier et al. (2014) for details). The secular component of the angular velocity time series arises because the surface mass load of ice on the continents, averaged in time, is nonzero. **Figure 11(b)** shows the polar wander path for the final cycle of loading and unloading for each of the example combinations of glaciation history and radial viscoelastic structure with the present day corresponding to the origin in the x–y plane. The 'loops' in the polar wander path are associated with meltwater pulses 1*A* and 1*B* that are clearly defined in the Barbados record of Peltier and Fairbanks (2006).

The time-domain form of the Liouville equation that connects these variations in the Cartesian components of the angular velocity vector of the planet to the corresponding elements of the products of inertia then states (e.g., Munk and MacDonald, 1960)

$$\frac{i\mathrm{d}\overline{m}/\mathrm{d}t}{\Omega\left(\frac{C-A}{A}\right)}+\overline{m}=\frac{1}{(C-A)}\left(\overline{I}-\frac{i\mathrm{d}\overline{I}/\mathrm{d}t}{\Omega}\right) \qquad [41]$$

In writing this equation, we have here employed the conventional complex variable-based definitions:

$$\begin{aligned}\overline{m}(t)&=m_x(t)+im_y(t)\\ \overline{I}(t)&=I_{xz}(t)+iI_{yz}(t)\end{aligned} \qquad [42]$$

In the low-frequency limit that is appropriate to understanding of the GIA process, eqn [35] reduces simply to the form

$$\overline{m}(t)=\frac{\overline{I}(t)}{(C-A)} \qquad [43]$$

In the expression for $\overline{I}$ in the preceding text, it is to be understood that this consists of two distinct parts, namely,

$$\overline{I}(t)=\overline{I}_{\mathrm{LOAD}}(t)+\overline{I}_{\mathrm{ROT}}(t) \qquad [44]$$

In **Figure 12**, we illustrate the variations of each of these contributions to the products of inertia that govern the polar wander effect for a single glacial cycle and the sum of these contributions according to eqn [38] along with a diagnostic constructed to test the accuracy of the mathematical methods employed to solve the low-frequency approximation to the Liouville equation [37]. The diagnostic is labeled 'ratio,' and the fact that it is unity demonstrates that the Liouville equation is being satisfied to high accuracy (see Peltier et al. (2012) for a detailed discussion). Given these models for a single ice-age cycle of the variation of I_{xz} and I_{yz} and a corresponding model for I_{zz} (see Peltier and Luthcke (2009) for an example of the latter), we are in a position to make detailed predictions of the time rate for change of the length of day and of polar wander speed and direction for a multicycle model of Late Quaternary glaciation and deglaciation, which we may compare to observations.

9.09.4 Observations of Millennial Scale Secular Variations in the Earth's Rotation Anomalies and the Impact of the Global Warming Era upon Them

Although the existence of a secular trend, persisting over thousands of years, in the nontidal variations of the l.o.d., has been known for at least the last quarter of the twentieth century and an equivalently important secular drift in the position of the pole has also become evident in the same period, debate has continued as to the cause of these secular variations in the Earth's rotation anomalies. Only through the development of the theory described in the last subsections of this chapter, however, has it proved possible to unambiguously attribute

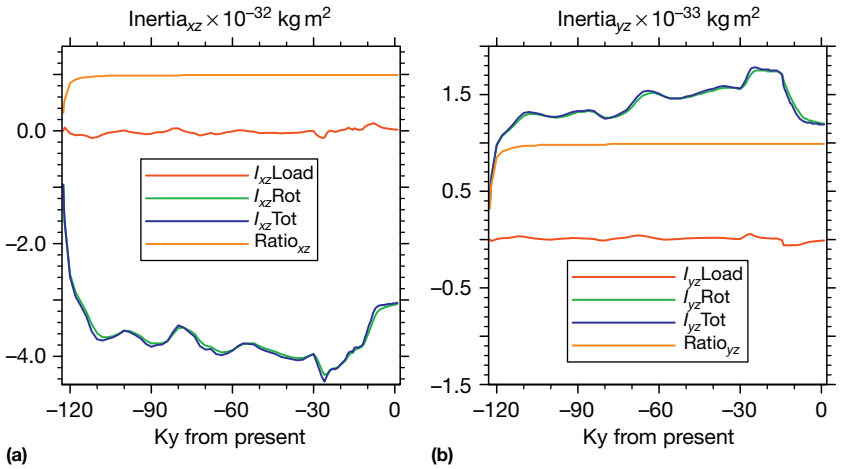


Figure 12 The primary components of the 'total' perturbations of the moment of inertia tensor that are responsible for driving the polar motion associated with the glaciation cycle shown in **Figure 11**. Also shown are the individual contributions to the total perturbation due to the surface mass 'load' and to the influence of rotational feedback, which is denoted 'rot.' The time series labeled ratio in each plate must equal unity if the combination of the time series of the inertia perturbations and the time series of variations in the components of the angular velocity satisfy the Liouville equation. The theory that has been developed to solve this equation is clearly highly accurate on this basis.

both signals to the GIA process. Prior to reviewing the analyses that have established this result, it will be useful to briefly review the observations themselves. As we will also demonstrate, these secular rates of change have themselves begun to deviate sharply from their previously stable rates over the past several decades. As we will argue, these sharp changes in the rates of secular variation are associated with the ongoing influence of the warming of the lower atmosphere that is due to the increasing concentration of carbon dioxide due to fossil fuel burning.

9.09.4.1 Ancient Eclipse Observations and the Nontidal Acceleration of Rotation

That an anomaly in Earth's rate of axial rotation did exist, beyond the relatively well-understood decrease due to the action of the tidal friction that arises because of the frictional dissipation of the tide raised in the Earth's oceans by (primarily) the gravitational attraction of the Moon, was first suggested (e.g., Morrison, 1973; Muller and Stephenson, 1975; Newton, 1972) on the basis of analyses of ancient eclipse data extending back more than 2500 years before present. For much of this interval of time, in both Babylon and China, naked-eye astronomers kept careful records of the timing and place of occurrence of total eclipses of the Sun and Moon. On the basis of the assumption that tidal friction had remained constant over this interval, a highly reasonable assumption as sea level has not changed appreciably over this Late Holocene period, one may accurately predict when and where a total eclipse of the Sun or Moon should have occurred. Analysis of such data, as described most recently in Stephenson and Morrison (1995) and Morrison and Stephenson (2001), clearly demonstrates that the further back in time for which an eclipse prediction is made, the larger is the error in timing, the sign of the error being such as to imply the action of a acceleration of rotation that is acting opposite to the effect of tidal friction. The compilation of data from the paper by Stephenson and Morrison (1995) is shown in **Figure 13**. Based upon the analysis of both lunar and solar eclipses that occurred between 700 BC and AD 1600, these authors reported an increase in the length of the mean solar day (l.o.d) of (1.7 ± 0.5) ms per year, which implies a rate of decrease of the angular speed of rotation of $(-4.5 \pm 0.1) \times 10^{-22}\,\mathrm{rad\,s^{-2}}$ on average over the past 2.7 ky. After subtracting the contribution due to tidal friction, which is accurately known based upon observations of the rate of recession of the Moon using Lunar Laser Ranging, Stephenson and Morrison (1995) inferred an average rate of nontidal acceleration of $(1.6 \pm 0.4) \times 10^{-22}\,\mathrm{rad\,s^{-2}}$ over this period. This value for the nontidal acceleration of rotation corresponds to a value for $\dot{J}_2$ of $(-3.5 \pm 0.8) \times 10^{-11}\,\mathrm{year^{-1}}$. As discussed in **Section 9.09.2**, this parameter provides a direct measure of the rate of change of the oblateness of planetary shape such that the larger J_2, the larger the oblateness or, equivalently, the polar moment of inertia. The fact that the time derivative is negative

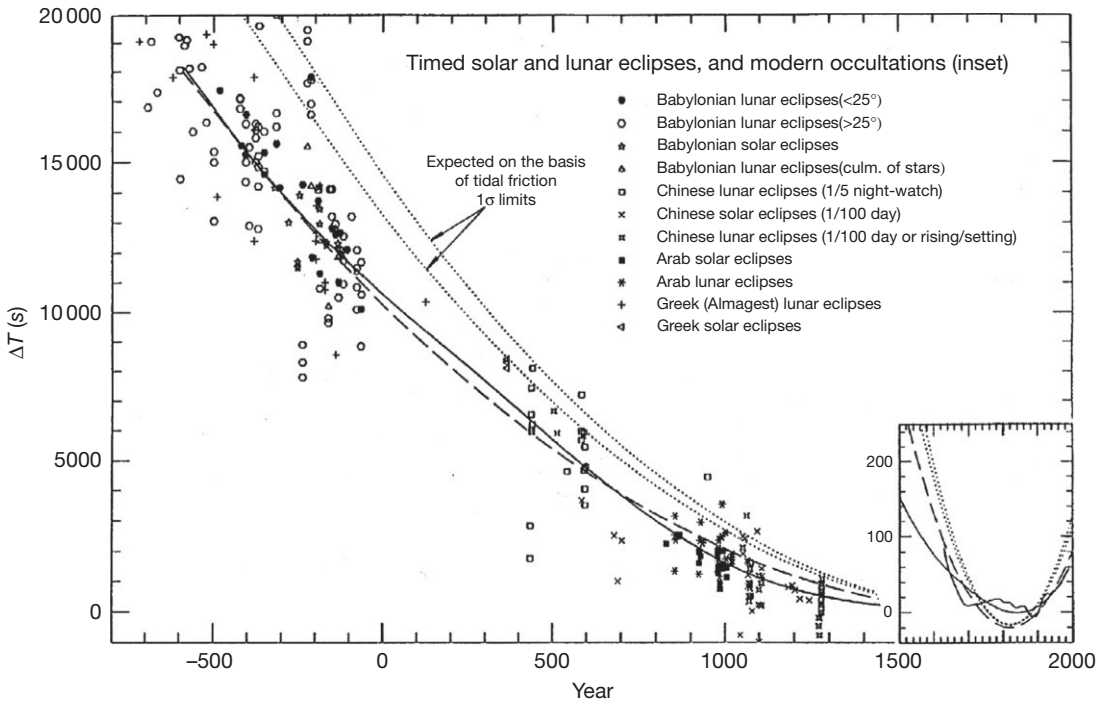


Figure 13 Results obtained for the time difference between the observed time of a total eclipse of the Sun or the Moon and the time at which the event was predicted to occur based upon the assumption of constant tidal torque. The inset (on a scale 25 times greater) includes the continuous curve derived from the lunar occultations in the period AD 1620–1955.5 and the difference TAI-UT1 + 32.1845 from 1955.5 to 1990. The dashed curve is the best fitting parabola $31t^2s$. The solid curve is fitted using cubic splines. This curve, the best fitting parabola, and the parabola expected on the basis of tidal friction are continued on the inset. Reproduced from Stephenson ER and Morrison LV (1995) Long term fluctuations in the earth's rotation: 700 B.C. to A.D. 1990. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences A* 351: 165–202.

therefore implies that the nontidal acceleration of rotation is derivative of a secular decrease in the oblateness of figure and therefore of a decrease in the value of the polar moment of inertia. That this is plausibly a consequence of the GIA process follows from the fact that at glacial maximum, the oblateness would have been increased by the increased surface mass load at the poles. Once this load is eliminated during the deglaciation process, this ice-age contribution to enhanced oblateness would have begun to relax viscoelastically as the planet returned to the more spherical shape characteristic of an interglacial period. Since angular momentum is conserved in the absence of external torques, the rate of axial rotation would have continued to increase thus leading to the observed nontidal *acceleration* of rotation.

Beginning with the first publication and interpretation of data from the Laser Geodynamics Satellite (LAGEOS) by Yoder et al. (1983), which appeared simultaneously with the interpretation of the nontidal acceleration as a consequence of the GIA effect (Peltier, 1982, 1983), the validity of the ancient eclipse-based inference of the nontidal acceleration was clearly established. Many estimates of the nontidal acceleration based upon the analysis of SLR data have now appeared (the effect observed is an acceleration in the rate of precession of the node of the orbit of the satellite). A relatively complete set of published estimates up to 1991 is shown in **Figure 14** where the estimates of various authors are provided in terms of $\dot{J}_2$. As will become clear in what follows, my reason for truncating the list of available estimates at 1991 has to do with the fact that subsequent to this time, a pronounced shift in the value of this parameter has become evident.

As a final comment upon the most recent research that has been conducted on l.o.d changes on millennial timescale, it is worth noting the work of Dunberry and Bloxham (2006) on the possibility of the existence of an oscillation in the l.o.d. that may be evident in the historical eclipse data set, an oscillation that the authors suggest may have a period of 1500 years. A similar timescale is suggested to be evident from archaeomagnetic artifacts, lake sediments, and lava flows (Daly and Le Goff, 1996; Hongre et al., 1998; Korte and Constable, 2003). They interpret this as plausibly arising as a consequence of coupling between the outer core and the mantle across the core–mantle interface. It should be noted, however, that the existence of any such 1500-year oscillatory component in the l.o.d. variations has been questioned by Dalmau (1997).

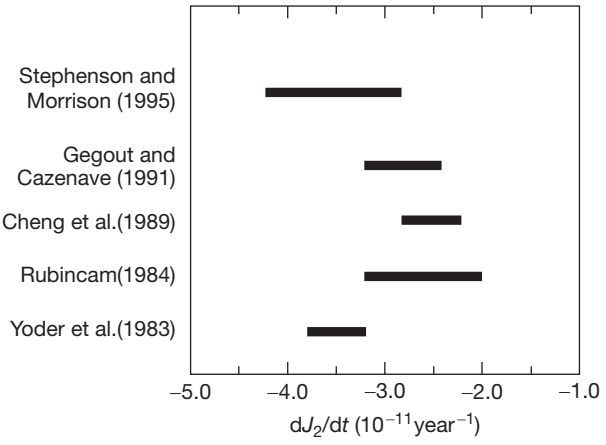


Figure 14 Published estimates of the parameter $\dot{J}_2$.

9.09.4.2 Millenial Timescale Polar Wander and the Glaciation–Deglaciation Cycle

Insofar as the second of the secular anomalies in the Earth's rotation is concerned, namely, the true wander of the north pole of rotation relative to the surface geography, the data originally obtained by the International Latitude Service that have been employed to constrain it are shown in **Figure 15**. The data initially tabulated by Vincente and Yumi (1969, 1970) were derived on the basis of observations of star transits using a northern-hemisphere array of photo-zenith tubes. The data consist of time series of the x and y coordinates of the position of the north pole of rotation relative to a reference frame fixed to the Earth's crust with origin at the CIO. This reference frame is shown in the inset polar projection in **Figure 15** and has its x-axis aligned with the Greenwich meridian. Inspection of the x and y time series demonstrates them to be dominated by an oscillatory pattern of beats with a period of $\sim$7 years. This pattern is well understood to arise as a consequence of the interference between the 12-month periodic annual wobble and the $\sim$14-month periodic Chandler wobble that were discussed in **Section 9.09.2** of this chapter. Based upon the review of our understanding of these features of the Earth's rotational history discussed therein, it would appear that previous controversies concerning their excitation have now been satisfactorily resolved. According to Dickman (1977), the true polar wander (TPW) evident as the secular drift of the position of the pole upon which the oscillatory signal is superimposed was occurring at a rate of $(0.95 \pm 0.15)^\circ\,\mathrm{My}^{-1}$ along the 75.5° W meridian over this time period as shown in the inset polar projection in **Figure 15**. An unambiguous connection of this anomaly of the Earth's rotation to the GIA process was first suggested in Peltier (1982) and Wu and Peltier (1984), whose detailed theoretical analysis, since further generalized in Peltier and Jiang (1996), showed that the initial analysis by Munk and MacDonald (1960) of the excitation of polar wander by surface loading of a viscoelastic model of the Earth did not in fact rule out an explanation in terms of the GIA process even though the surface mass load during the Late Holocene interglacial period has been very nearly time-invariant. The analysis of the data by Dickman (1977) was preceded by several earlier analyses and has been followed by others. Most of these results are shown in **Figure 16** where they are accompanied by an early result derived on the basis of the analysis of VLBI observations by Carter et al. (1986).

The most recent analysis of the ILS data is that of Argus and Gross (2004), who have investigated the possibility that all of these interpretations may be contaminated by the failure to account for the influence of surface plate tectonic motions. Their analysis of the raw data delivered the best estimate of the speed of polar wander of $1.06^\circ\,\mathrm{My}^{-1}$. Depending upon how this estimate was corrected for the influence of plate tectonics, however, considerably different results could be obtained, as listed on **Table 1** of that paper. Whereas the ILS inference uncorrected for plate motion was that the ongoing polar wander was along the 75.5° W meridian, if the same data set is corrected by making the inference in the frame of

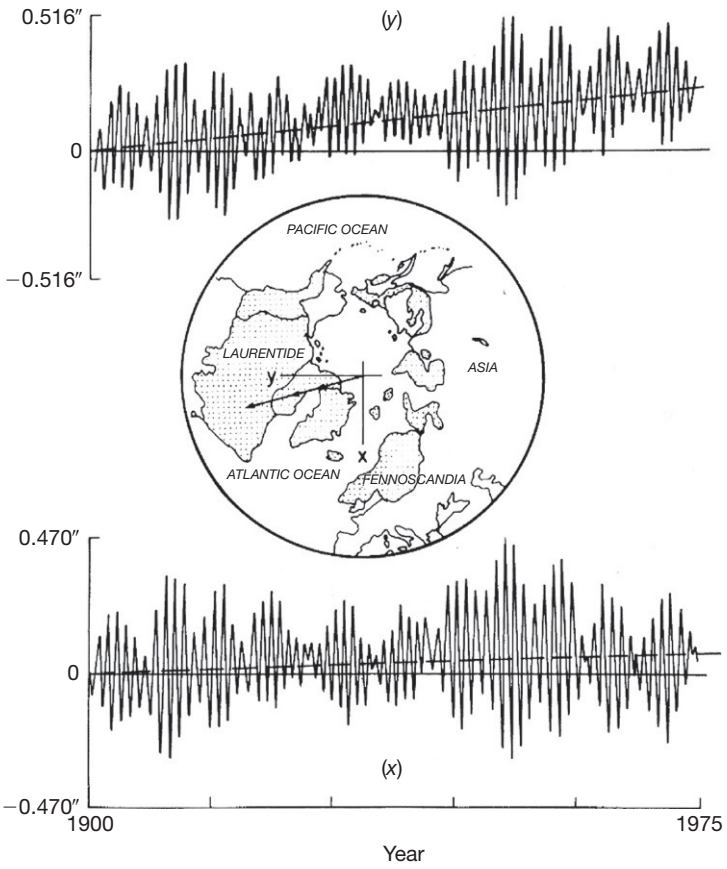


Figure 15 Time series of the location of the north pole of rotation of the planet, as *x* and *y* coordinates relative to the coordinate system shown in the inset polar projection with origin at the Conventional International Origin (CIO). These data are those of the International Latitude Service (Vincente and Yumi, 1969, 1970), and they reveal a secular drift of the pole at a rate near 0.95° My^{-1} approximately along the 75.5° W meridian of longitude.

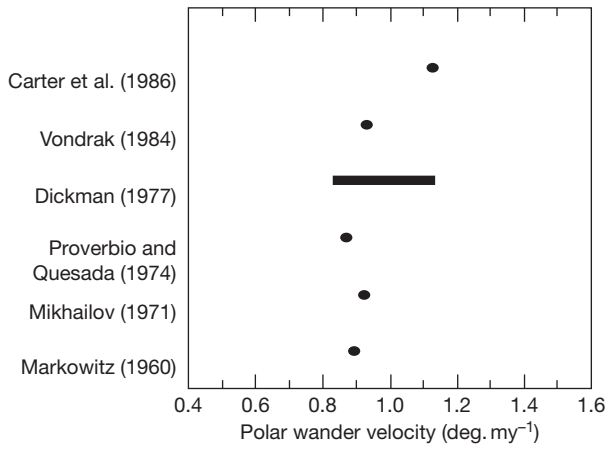


Figure 16 Published estimates of the speed of polar motion based primarily upon the ILS data.

reference in which the lithosphere exhibits no net rotation, then the corresponding speed and direction change slightly to 0.98° My^{-1} southward along the 79.9° W meridian. However, if the correction to these data is based upon the 'hot spot frame,' then one obtains, from the ILS data, according to Argus and Gross (2004), the values 1.12° My^{-1} for the speed and southward along the 69° W meridian for the direction, a significant difference. The fact that all of these inferences of polar wander direction suggest it to be moving 'towards' what was the centroid of the ancient Laurentide ice sheet (see Figure 7) is nevertheless highly significant insofar as the explanation in terms of glacial isostasy is concerned. Since the constraint of angular momentum conservation requires that the pole must drift so as to align itself along the axis relative to which the moment of inertia is largest, it is clear why, as a consequence of the GIA effect, the pole should now be moving towards what was the centroid of the greatest concentration of land ice. Following deglaciation, there exists a residual depression of the land that has yet to be eliminated by the slow viscous process of GIA. This region therefore constitutes a localized deficit of mass relative to the state of gravitational equilibrium that will eventually be reached once the GIA-related relaxation of shape has been completed. It is towards this regional mass deficit that the pole must move in order to come into alignment with the axis of greatest inertia. That it never reaches and in general will not reach this region is simply a consequence of the fact that the speed of polar wander decreases as a function of time as the shape relaxation proceeds at a rate determined by mantle viscosity.

Table 1 Interstudy comparison of secular trends in polar wander and $\dot{J}_2$

Source	*Fits*	*Time covered*	*Polar motion (mas year^{-1})*	*Direction*
Polar wander				
Argus and Gross (2004) (ILS) (relative to mean lithosphere)	1	1899–1979	3.53	79.9° W
Gross and Vondrák (1999) (HIPPARCOS)	1	1900–1992	3.51	79.2° W
Gross and Vondrák (1999) (SPACE96)[a]	1	1976–1997	4.12	73.9° W
Gross and Poutanen (2009) (SPACE2007)[a]	2	1976–1995	4.23	72.2° W
		1995–2007	2.92	64.8° W
This study (SPACE2008)	2	1976–1992	4.5(±0.1)	68(±8)° W
		1992–2008	1.8(±0.4)	58(±9)° W

Source	*Fits*	*Time covered*	$\dot{J}_2$ ($\times 10^{-11}$ year^{-1})
$\dot{J}_2$			
Yoder et al. (1983)	1	1978–1983	−3.5
Stephenson and Morrison (1995)	1	700 BC–AD 1600	−3.5
Cheng and Tapley (2004)	1	1976–2004	−2.75
This study	2	1976–1992	−3.7 (±0.1)
		1992–2009	−0.9 (±0.2)

[a]SPACE2007 and SPACE96 are two earlier versions of the SPACE2008 Earth Orientation Parameters-based data series for polar motion.

Reproduced from Roy K and Peltier WR (2011) GRACE era secular trends in Earth rotation parameters: A global scale impact of the global warming process? *Geophysical Research Letters* 38: L10306, http://dx.doi.org/10.1029/2011GL047282.2011.

The question of the accuracy of the direction of the secular drift of the pole that is derived on the basis of the ILS data may be at least as important as the speed. The reason for this has to do with the use of the rotational anomalies to constrain the rate of melting of land ice that may be occurring at present due to the action of greenhouse gas-induced global warming. Depending upon the polar wander prediction due to the continuing action of the GIA effect, there will exist a residual between this prediction and the modern observations that may be employed to constrain the rate and geographic locations of modern sources of land-ice melting (Greenland, Antarctica, small ice sheets and glaciers, see Dyurgerov and Meier (2005)) as previously discussed in Peltier (e.g., 1998, see Figure 46 of that paper). Further comment on this issue will follow in the next section of this chapter in which the theory reviewed in **Section 9.09.3** will be applied to understanding the observational constraints upon pre-1990s polar wander described in this section. First, however, we will show here that a very substantial shift in the trends of both rotational anomalies has occurred over the past several decades.

9.09.4.3 Global Warming Era Shifts in the Millennial Timescale Trends in the Earth's Rotation Anomalies Based upon Modern Earth Orientation Data Sets

Modern trends in the polar wander signal have been examined using a recent set of independent space-geodetic Earth orientation measurements, which are generated annually by the Jet Propulsion Laboratory (JPL). The particular set analyzed in the recent paper by Roy and Peltier (2011), known as SPACE2008, was made available from JPL's Geodynamics and Space Geodesy Group via anonymous ftp at <ftp://euler.jpl.nasa.gov/keof/combinations/2008> and spanned the period from 28 September 1976 to 2 July 2009. It includes the values of the polar motion components at daily intervals, with components extending from the CIO along an x-axis aligned with the Greenwich meridian and along a y-axis aligned with the 90° W meridian (Ratcliff and Gross, 2010). As discussed earlier, polar motion is dominated by large high-frequency variations, and in order to study the linear trends in the series influenced by the process of GIA, it is imperative to remove those signals. In Roy and Peltier (2011), a Butterworth-type low-pass filter was applied to the polar motion series (Butterworth, 1930), with a period cutoff set at 6 years, a value that is consistent with previous studies of polar motion trends (Gross and Vondrák, 1999). The Butterworth filter is preferred to simpler, idealized digital filters such as the simple boxcar filter, as it has a maximally flat frequency response and phase shift effects that are simple to compensate. The smoothed series, shown in **Figure 17(a)** and **17(c)** (for the x- and y-components, respectively), served as the basis for the analysis in Roy and Peltier (2011) of quasiperiodic low-frequency residual variability, using a methodology similar to that employed in Gross and Vondrák (1999). Low-frequency and quasiperiodic terms were determined by performing a simultaneous weighted least-squares fit for a mean, a linear trend, and low-frequency periodic terms that correspond to the most prominent peaks in the amplitude spectrum of the time series. Changes in the linear trend were isolated by separating the time series into two subsets. The position of the 'pivot time' that separates the two adjacent periods of best fit was allowed to vary, and the linear trend and mean were adjusted as this 'knot' was shifted in time in order to find the pivot time that minimizes the overall root-mean-square error of the two-segment fit. This new fit is then compared with the best overall fit with a single linear trend and must be characterized by a substantially lower root-mean-square error of fit to be considered plausible physically. The resulting two-segment fits are presented in **Figure 17(b)** and **17(d)** for both the x- and y-components, respectively, and are listed in **Table 1**.

Recent variations in the linear trends for the nontidal acceleration of planetary rotation rate are studied using the most recently available satellite laser ranging (SLR) data for a subset of eight geodetic satellites (Starlette, Ajisai, Stella, LAGEOS 1 and 2,

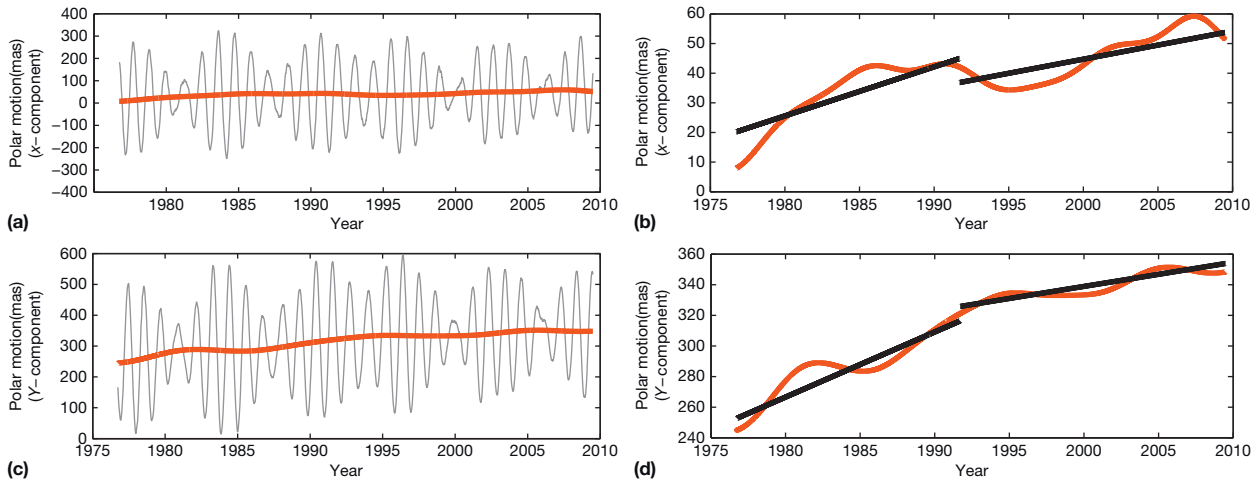


Figure 17 (a) and (b) show the *x*- and *y*-components of the polar motion for the period ~1975–2010 based upon the Earth orientation data set discussed in the text. (c) and (d) show both the Butterworth low-pass filtered versions of the time series in (a) and (b) together with two-segment linear fits designed to optimally reduce the RMS misfit of the secular rate(s) of change compared with that determined by a single-segment fit. The key result, which is from the paper by Roy and Peltier (2011), is that a marked shift in the secular rate of change occurred in the early 1990s.

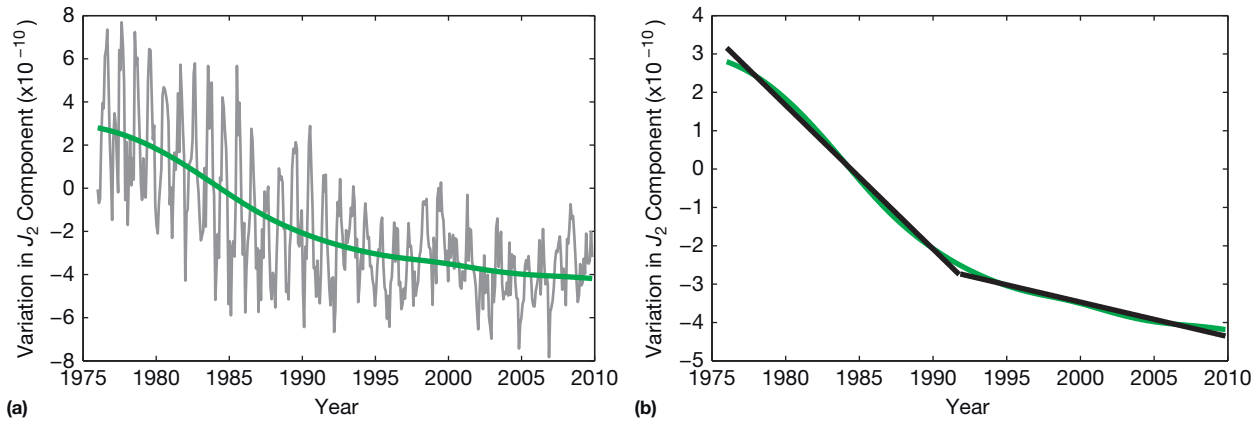


Figure 18 The results of a similar analysis to that shown in **Figure 17** but here for the time series of J_2. (a) shows the raw time series data and the Butterworth low-pass filtered version, while (b) shows the two-segment fit to the low-pass filtered time series in (a), which once more establishes the timing of the apparent shift in the secular variation of the time series as having occurred in the early 1990s. This result is again from Roy and Peltier (2011).

Etalon-1 and Etalon-2, and BE-C (from 1999 onward)) (Cheng and Tapley, 2004; Tapley et al., 2004). The available data, obtained originally via anonymous ftp at <ftp://ftp.csr.utexas.edu/outgoing/cheng/j2.76010911>, span January 1976 to the end of 2009 and are presented in **Figure 18(a)**, also from the study of Roy and Peltier, 2011. In order to study variations in the linear trend over this period, low-frequency terms are also eliminated using a low-pass Butterworth filter. The cutoff in this case was fixed at 20 years, which is sufficient to remove suggested 18.6-year tidal and decadal variations and is consistent with the analysis of Cheng and Tapley (2004). The smoothed series, plotted in **Figure 18(a)**, reveals a marked change in approximately the mid-1990s. Linear trends were then determined in the smoothed J_2 series, by separating the data series into two segments and fitting the trend separately for each part, as for the polar motion. The pivot time was again shifted in time to find the position of the 'knot' in time that minimizes the overall root-mean-square error of fit, the final result of which is presented in **Figure 18(b)**.

Changes in the recent linear trends for the secular drift of the position of the pole relative to the surface geography and the nontidal acceleration of the rate of planetary rotation have been independently revealed to exist through these analyses. However, it is most instructive to compare the variation of the total root-mean-square error for the two rotation-related time series as the position of the transition in the linear trend is allowed to vary. These results, which are presented in **Figure 19**, show that the transition in the linear trends for both anomalies occurs near 1992.

As shown in **Table 1**, this study of the observations of the polar motion time series suggests that its rate remained approximately fixed, prior to 1992, at $4.5(\pm 0.1)$ mas year^{-1} ($1.3°\,\mathrm{My}^{-1}$), along the $68(\pm 8)°$W meridian, while it slows to $1.8(\pm 0.4)$ mas year^{-1} ($0.5°\,\mathrm{My}^{-1}$) after 1992, along the $58(\pm 9)°$W meridian. Although not corrected for plate tectonics, as in Argus and Gross (2004), the Roy and Peltier (2011) analysis demonstrates that dividing the time series into two distinct epochs increases the magnitude of the established

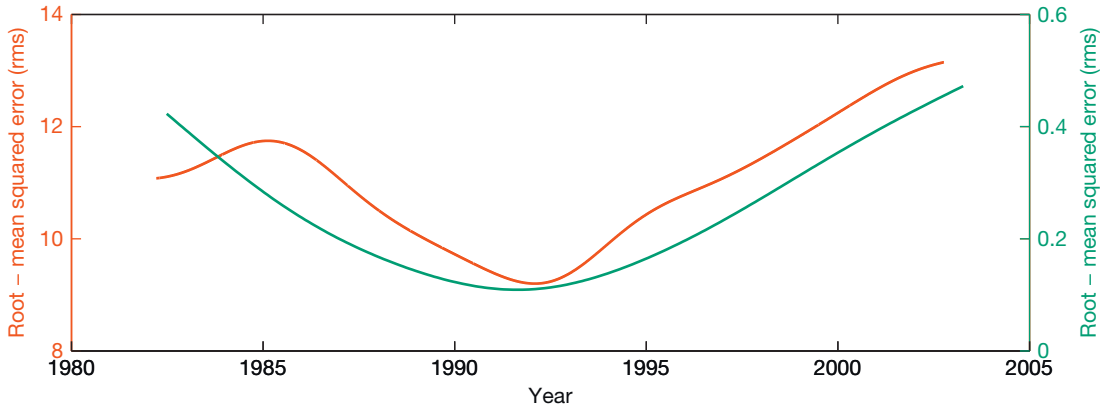


Figure 19 The RMS misfit of the two-linear-segment fits relative to that of a single-segment fit for both the polar motion and length-of-day related observables as a function of the time selected for the pivot point in time at which the secular variation is assumed to change. The minimum in the misfit for both time series is thereby shown to occur in ~1992, a full decade prior to the launch of the GRACE satellites.

secular rates of change prior to 1992, but decreases them for modern times subsequent to this. With regard to changes in the J_2 component, the Roy–Peltier analyses suggest a rate of change of $-3.7(\pm0.1)\times10^{-11}$ year^{-1} before 1992 and $-0.9(\pm0.2)\times10^{-11}$ year^{-1} after 1992. Dividing the data series into two parts is very significant, because it better reconciles the observed rate of change for J_2 presented in Cheng and Tapley (2004) with the estimate from historical records of Stephenson and Morrison (1995) ($-3.5(\pm0.8)\times10^{-11}$ year^{-1}).

The inference of a notable change in the linear trends for both anomalies is highly significant, as both the variation in J_2 and the polar wander data are dependent on completely independent elements of the Earth's moment of inertia tensor (e.g., Peltier and Luthcke, 2009). Models of the GIA process, in which the moment of inertia tensor of the planet is determined by a specific model of the time-dependent continental ice-sheet loading and of the depth dependence of mantle viscosity, are able to fit simultaneously those two rotational anomalies with great success, which suggests that it would be very surprising if the observed change in the linear trends inferred to have occurred since 1992 was not caused by recent changes in the surface ice-sheet loading on the planet.

The existence of a link between the timing of the changes in the linear trends for the rotational state anomalies and recent changes in the ice-sheet loading on the planet is supported by studies of the year-to-year melting rates in major continental ice sheets around the world. Thomas et al. (2006), for example, compared various diverging estimates of elevation changes and mass balance for the Greenland ice sheet and suggested that satellite altimetry measurements might seriously underestimate Greenland ice losses. If that were the case, all other indicators, such as snow-accumulation rates and airborne laser altimetry measurements, support the idea of a sharp decrease in ice-sheet mass starting in the mid-1990s.

9.09.5 Earth's Rotational Response to the Cyclic Glaciation Cycle of Late Pleistocene Time: Data–Model Comparisons

Based upon the discussion in Section 9.09.3 of this chapter, it will be clear that accurate predictions of relative sea-level history are a required preliminary to accurate predictions of the rotational response to the GIA process and thus to the understanding of the Earth's rotation history on millennial timescales. Since the best test of the accuracy of relative sea-level history predictions based upon solution of the SLE is provided by radiocarbon-dated relative sea-level curves from coastal locations, it will be clear that the confrontation of theory with observational 'truth' requires a voluminous compilation of such records, individual examples of which have been produced by innumerable scientists working in the area of Pleistocene geomorphology over the past half century. That a global database of such records is expected to record a wide range of signatures of the GIA process will be clear on the basis of Figure 20, which shows predictions of the present-day rate of relative sea-level rise for the ICE-6G (VM5a) model both including Figure 20(a) and excluding 20(b) the influence of rotational feedback. The difference between these predictions (Figure 20(a) and 20(b)) is shown in Figure 20(c), which establishes the dominant role that the polar wander component of the rotational response to the GIA effect plays in contributing to sea-level history. The 'signature' of this feedback is the existence of the strong degree 2 and order 1 spherical harmonic component, a quadrupolar distribution that is the expected form of this feedback based upon eqns [13] in Section 9.09.3. Figure 20(d) shows the sum of the signal shown in Figure 20(a) for the present-day rate of relative sea-level rise and the prediction for the same model of the present-day rate of radial displacement of the crust. As first discussed in Peltier (1999), this sum represents the time rate of change of geoid height due to the GIA process, an approximation to which is being measured by the Gravity Recovery and Climate Experiment (GRACE) dual-satellite system (see Figure 21) that is now in space. The reason why GRACE is able to observe only an approximation to the actual sea level-defined geoid has been discussed in detail in the recent paper of Peltier et al. (2012). In order to test the accuracy of predictions of the Earth's rotational response to the GIA process produced by the theory discussed in Section 9.09.4, an appropriate strategy is therefore to focus upon the observational constraints provided by geologic recordings of postglacial relative sea-level change through the Holocene interval of time in those regions in which the impact of rotational feedback is expected to be most intense.

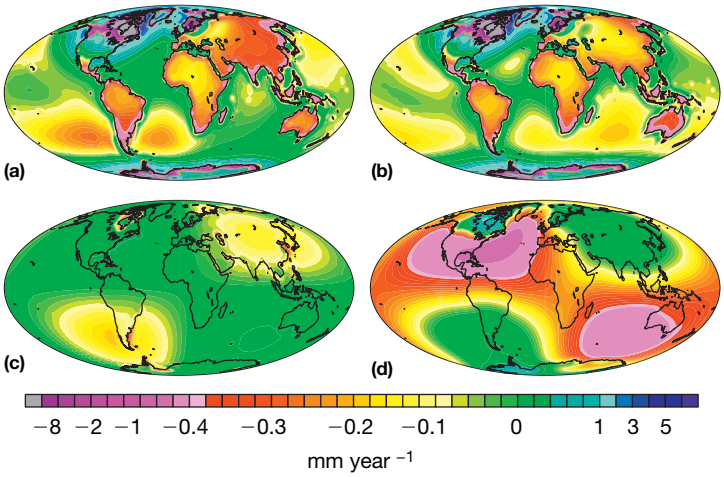


Figure 20 (a) The ICE-6G (VM5a) model prediction of (a) the present-day rate of sea-level rise relative to the deforming surface of the solid Earth, including the influence of rotational feedback; (b) same as (a) but excluding the influence of rotational feedback; (c) the difference between (a) and (b) illustrating the spherical harmonic degree 2 and order 1 pattern that dominates the influence due to rotational feedback on account of the control exerted by the polar wander effect; (d) the present-day geoid height time dependence predicted by the ICE-6G (VM5a) model, obtained by adding to the field in (a) the prediction of the time rate of change of radial displacement of the surface of the solid Earth with respect to the center of mass of the planet.

9.09.5.1 A Database of Holocene Relative Sea-Level Histories

A location map of the positions on Earth's surface from which ^{14}C-dated relative sea-level histories are available in the University of Toronto database is shown in **Figure 22** (from Peltier (1998)). Comparing the coverage provided by these many hundreds of individual RSL curves with the geographically intricate pattern evident in **Figure 20**, it will be clear that there exists, not surprisingly, much better coverage of this signature of the GIA process in the northern hemisphere than is available from the southern hemisphere. Yet, as will be made clear in what follows, the role of the available southern hemisphere data in confirming the important role that variations in the Earth's rotation play in forming the detailed signature of the GIA effect evident in Holocene RSL records is extremely important. **Figure 23** compares a number of ^{14}C-dated RSL records from this database with the predictions of the ICE-5G (VM2) and ICE-6G (VM5a) models that served as the basis for the computation of the time series of the variations in the elements of the moment of inertia tensor that are required in the computation of the rotational response to the glaciation–deglaciation cycle. The locations of the individual sites are shown in the map of the prediction of the present-day rate of sea-level rise predicted by the ICE-6G (VM5a) model of the GIA process that forms the central portion of **Figure 23**. Inspection of this initial set of comparisons will show that the theoretical structure developed to explain such widely varying histories of postglacial sea-level change is able to capture the details of this variability in a highly accurate fashion.

9.09.5.2 The Influence of Rotational Feedback upon Postglacial Relative Sea-Level History and Its Impact upon Predictions of the Earth's Rotation Anomalies

Although the purpose of this subsection of this chapter will be to explore the extent to which the most recent models of the GIA process may be employed to further reinforce the interpretation of the millennial timescale variations in the Earth's rotation anomalies in terms of the GIA process, it will be useful to begin by revisiting the results previously obtained by Peltier and Jiang (1996) who employed the ICE-4G deglaciation model of Peltier (1994, 1996b) to investigate the extent to which both of the previously discussed anomalies were explicable in these terms. Their results are summarized in **Figure 24** (reproduced from their paper), which shows, respectively, in **Figure 24(a)** and **24(b)** the theoretical predictions of $\dot{J}_2$ and polar wander speed as a function of the viscosity of the lower mantle of the Earth below a depth of 660 km, with the value of the viscosity of the upper mantle fixed to the nominally accurate value of 10^{21} Pa s, and the thickness of the surface lithosphere taken to be $L = 120$ km. Inspection of this Figure demonstrates that both of the millennial timescale secular anomalies are fit by the same simple model of the radial variation of mantle viscosity. With the upper-mantle viscosity fixed to the aforementioned value, the lower-mantle viscosity required to fit the observations is $\sim 2 \times 10^{21}$ Pa s. It might be seen as unlikely, given that the two rotational anomalies depend upon entirely independent elements of the moment of inertia tensor that this degree of agreement would exist if GIA were not the primary explanation for both observations.

It is extremely important, however, that we demonstrate that this result is robust against explicit incorporation of the difference between k_f and $k_2^T(s=0)$ as previously discussed in **Section 9.09.3**. This issue is explored in detail via **Figure 25**, which is from the recent paper of Peltier and Luthcke (2009), and in greater detail in **Figure 26**, also reproduced from their paper. These analyses are based upon the use of the ICE-5G loading history with seven cycles of glaciation and deglaciation included in the calculation. **Figure 25** may be compared directly with **Figure 24** and is shown to quite accurately reproduce that result for polar wander speed in the case in which the parameter $\Delta = 0$, which represents the EQM case in which

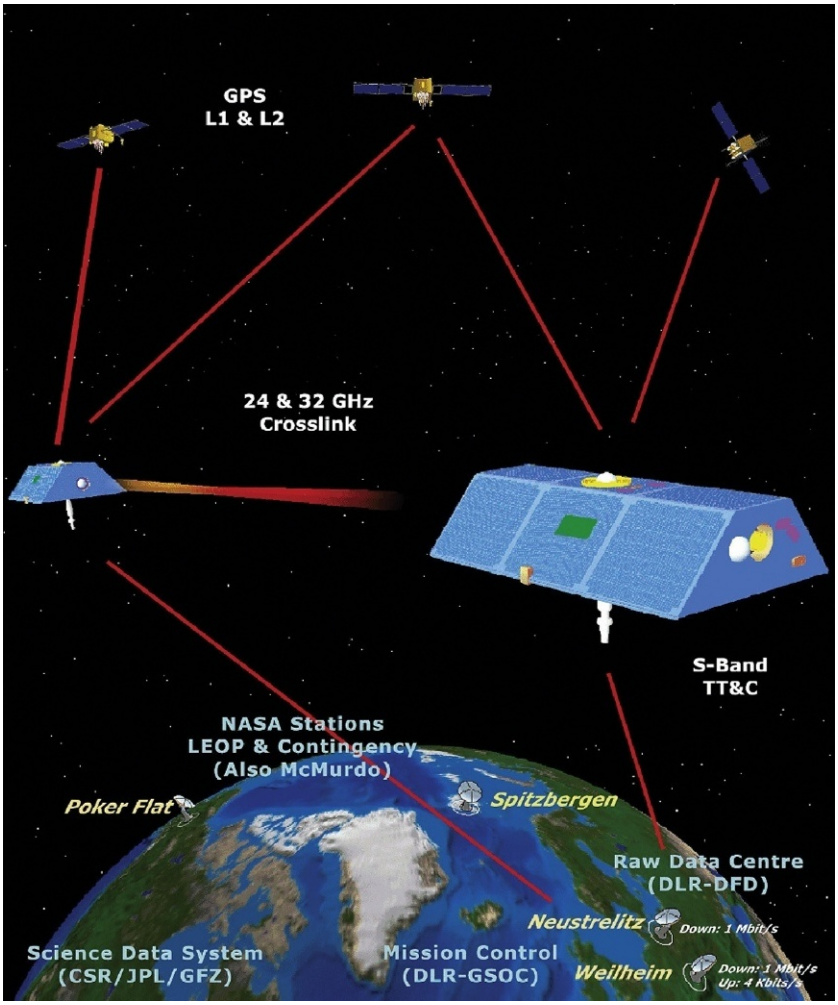


Figure 21 Schematic of the dual-satellite Gravity Recovery and Climate Experiment (GRACE) satellite experiment designed to measure the time dependence of the gravitational field of the planet. It is important to keep in mind that questions remain concerning the accuracy with which this system is able to directly measure the time-dependent amplitudes of the Stokes coefficients of degree 2 and order 1. These are most accurately obtained on the basis of the polar wander measurements of Roy and Peltier (2011) using the theoretical connections between these data and the crucial Stokes coefficients.

$k_f = k_2^T(s=0)$. On this figure, the phrase 'with rotation' is employed to represent the fact that the full impact of the change in ocean loading due to the changing rotation of the Earth is taken into account in the calculation. The case $\Delta = 1$ corresponds to the case in which not only is k_f different from $k_2^T(s=0)$ but also the elastic lithospheric thickness in the limit of infinite time is taken to be equal to 90 km, the same thickness as is approximately characteristic for short timescale continental ice-sheet cycle loading events. The two additional cases for which results are shown in the figure for polar wander speed correspond to cases in which the thickness of the lithosphere is taken to vanish in the infinite time limit but in which either one of the two plausible values of k_f shown in Figure 6 is assumed so that there continues to exist a difference between k_f and $k_2^T(s=0)$. Inspection of this figure demonstrates that the assumption of a finite value of lithospheric thickness in the infinite time limit leads to the failure of the model to simultaneously explain both of the pre-1992 values of the rotational anomalies. However, when the lithospheric thickness is assumed to vanish in the limit of infinite time, then, even when the difference between k_f and $k_2(s=0)$ is taken fully into account, the model still reasonably explains both data simultaneously.

In Figure 26, the robustness of the results shown in Figure 25 is explored by investigation of the sensitivity of the polar wander speed analyses to variations in the assumed value of the viscosity of the upper mantle and transition zone viscosity and to variations in the accuracy with which the representation of the ocean component of the surface mass loading cycle is included. Since the latest inferences of the mean viscosity of the upper mantle and transition zone fix this parameter to a value close to 0.4×10^{21} Pa s, the analyses demonstrate that this shift in value from the previously assumed value of 10^{21} Pa s has only a minor effect. However, the figure also suggests that it is important that the full influence of rotational feedback on the redistribution of mass in the global ocean is accurately calculated.

Since the aforementioned publication described two-layer viscosity model results of Peltier and Jiang (1996) and Peltier

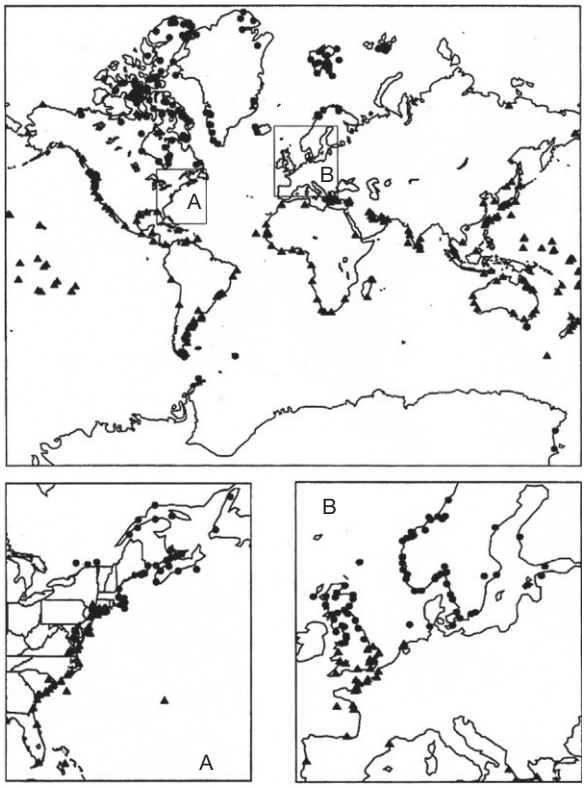


Figure 22 Global location map of the sites from which ^{14}C-dated time series of relative sea-level history are available in the University of Toronto database.

and Luthcke (2009), further refinements of the required analyses have been performed. In particular, in Peltier (1996b), the VM2 model of the radial viscosity structure was inferred on the basis of a full Bayesian inversion of the available data (see Figure 27). Equally important has been the considerable further refinement of the loading history that is represented by the ICE-5G model of Peltier (2004) and previously introduced in Section 9.09.3 of this chapter and the even more recent model ICE-6G (VM5a). The first row of Table 2 lists the predictions of present-day polar wander speed and direction and $\dot{J}_2$ for the ICE-6G (VM5aD1) model, assuming the validity of this model of the radial viscosity structure discussed in Peltier and Drummond (2010), which is a relatively minor variant upon the VM5a structure that includes a slight softening of the viscosity of the D″ layer above the core–mantle boundary. Model VM5a is itself a simple multilayer fit to VM2, but one that allows for viscosity stratification of the lower lithosphere. These analyses include the full influence of rotational feedback as discussed in Peltier et al. (2012) and demonstrate that the model well fits the pre-1992 values of all three of the rotational anomalies. The additional columns of Table 2 contain values of the degree 2 and order 1 Stokes coefficients in the spherical harmonic expansion of the time dependence of geoid height, a field that has been measured approximately by the GRACE satellite system that has been in space only since March of 2002 and therefore could not provide measurements of these parameters in the pre-1992 period subsequent to which a marked change in the Earth's rotation anomalies took place.

In the second row of Table 2, predictions of the same Earth's rotation observables for a model labeled 'A' are shown

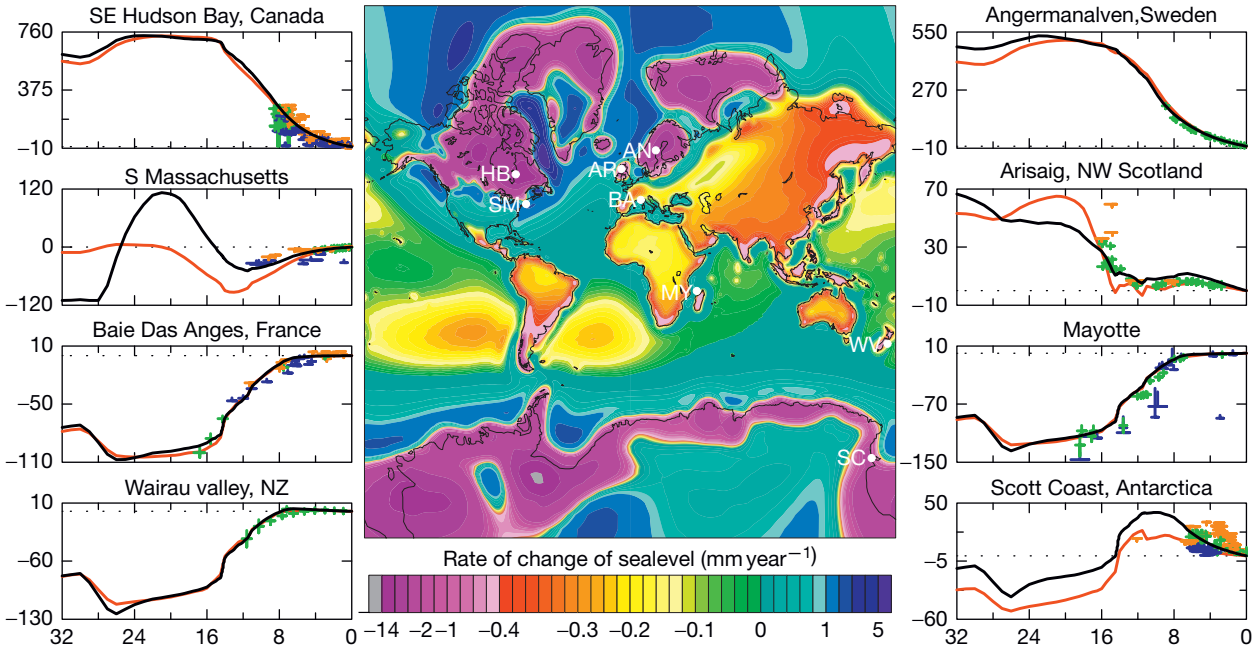


Figure 23 Illustrative suite of predictions of the history of relative sea-level change at eight sample locations using the ICE-5G(VM2) model of the global GIA process. The locations of the individual sites are superimposed upon a Mercator projection of the predicted present-day rate of relative sea-level rise. Predictions are shown for two slightly different versions of the model that differ from one another only in terms of the assumption made concerning the timing of Last Glacial Maximum. For one version of the model, LGM is assumed to have occurred at the conventional age of 21 ka; for the other, it is assumed, as illustrated in Figure 8, that LGM occurred considerably earlier at 26 ka as suggested on the basis of the Barbados data discussed in Peltier and Fairbanks (2006). The results for the ICE-6G (VM5a) model at these sites are very similar.

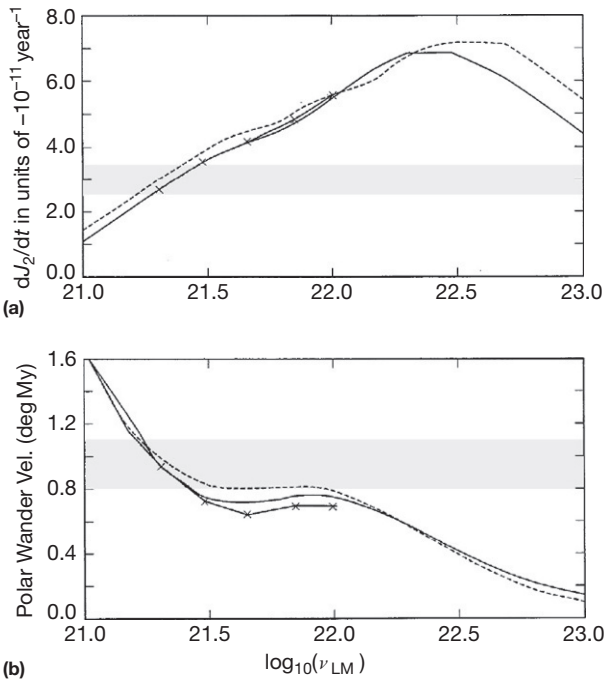


Figure 24 (a) $\dot{J}_2$ as a function of lower-mantle viscosity with the upper-mantle viscosity held fixed to the value of 10^{21} Pa s. (b) Same as (a) but for polar wander speed. In each frame, the hatched region represents the observationally constrained range based primarily upon data from the pre-1992 period. Note that the millennium timescale secular anomalies in the Earth's rotation are 'explained' as a consequence of the glacial isostatic adjustment process by the same model of the radial variation of mantle viscosity. Reproduced from Peltier WR and Jiang X (1996) Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *Journal of Geophysical Research* 101: 3269–3290 (Correction, *Journal of Geophysical Research* 102: 10101–10103, 1997).

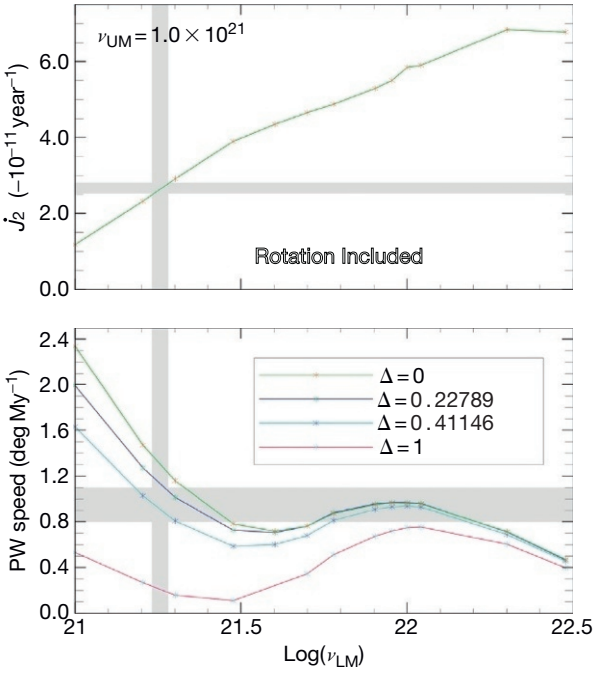


Figure 25 The earlier analysis shown in **Figure 24** as a function of the extent to which a difference between k_f and $k_2^T(s=0)$ is taken into account and is taken from the paper of Peltier and Luthcke (2009). The case $\Delta=0$ corresponds to the equivalent Earth model approach of Munk and MacDonald (1960) for which $k_f=k_2^T(s=0)$. The case $\Delta=1$ corresponds to a case in which the elastic lithosphere is assumed to maintain its global integrity in the limit $t\rightarrow\infty$. The additional cases both correspond to cases in which the infinite time limit of the elastic lithospheric thickness vanishes and in which either one or the other of the values of k_f shown in **Figure 7** is assumed. Notable is the fact that the finite infinite time elastic lithospheric thickness case is ruled out, whereas the zero infinite time thickness results continue to reconcile the data.

in which, aside from the surface load forcing associated with the Late Quaternary ice-age cycle, additional surface load forcing has been added to the model representing the modern global warming-induced land-ice melting that is currently ongoing in Greenland, Antarctica, and Alaska and that is directly constrained by analysis of the time-dependent gravity observations being provided by the GRACE satellite system. These GRACE-inferred melt rates have recently been reviewed in Peltier (2009) whose original analyses have confirmed the validity of the earlier results by Velicogna and Wahr (2006a,b) for Greenland and Antarctica but significantly reduced the inferred rate of melting from the high mountains of Alaska according to earlier analyses. The results from model A for $\dot{J}_2$ demonstrate that the influence of these additional sources of land-ice melting causes a marked reduction in $\dot{J}_2$ that appears to plausibly explain the strong reduction in the value of this parameter described in Roy and Peltier (2011). This is confirmed in **Figure 28** in the construction of which it is assumed that the base model accurately fits the pre-1992 observations and the anomalies associated with the melting of ice from these three polar ice-sheet catchments are simply added to the prediction of the base model. Although this additional forcing alone appears to fit the post-1992 shift in the value of $\dot{J}_2$ inferred by Roy and Peltier (2011), we will see in what follows that this explanation is seriously flawed. Not only does it not simultaneously fit the observed shift in polar wander speed and direction as will be clear by inspection of the data in the table but also critical features of the geoid height time dependence represented by the Stokes coefficients of degree 2 and order 1 are also violated.

9.09.5.3 Measurements of the Expected Quadrupolar 'Signature' of Rotational Feedback in Postglacial Sea-Level Histories

In order to adequately test the veracity of the theory that has been developed to compute the Earth's rotational response to the glaciation–deglaciation cycle and thereby to adequately test our understanding of the history of Earth's rotation on millennial timescales, the use of the global patterns of postglacial relative sea-level change predicted by the different versions of the calculation might be expected to be especially useful. As previously noted, these patterns are best illustrated by global predictions of the present-day rate of relative sea-level rise that would be expected to be observed as a secular rate of change on a tide gauge installed in a coastal location if the only contribution to relative sea-level history were the continuing impact of

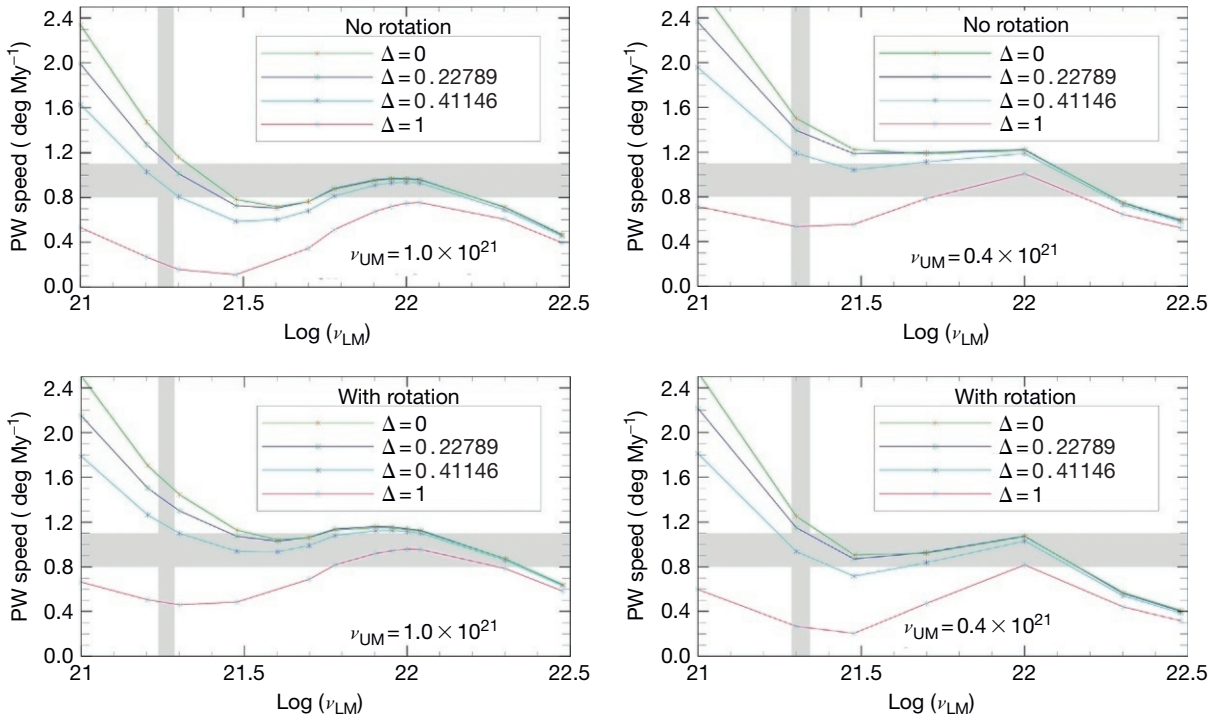


Figure 26 The results of analyses similar to those shown in the preceding text in **Figure 25** but solely for the polar wander speed datum. These results describe the sensitivity of the fits to the observation of polar wander speed in the pre-1990s era to two variations upon the basic calculation. These variations compare the sensitivity of the results to variations in the upper-mantle viscosity by changing the upper-mantle value from 1E21 to 0.4E21 Pa s and also to variations in the ocean component of the surface load by either including (with rotation) or excluding (no rotation) the redistribution of the water among and within the ocean basins due to the influence of rotational feedback.

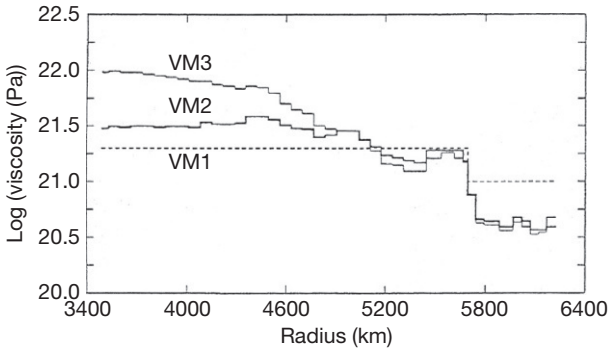


Figure 27 Viscosity models VM1, VM2, and VM3 originally inferred in Peltier (1994, 1996a,b,c) on the basis of full Bayesian inversions of a subset of the observational constraints pertaining to the GIA process. Model VM1 was the starting model employed to initiate the Bayesian inversion that employed the methodology of Tarantola and Valette (1982a,b). Control on the viscosity in the deepest mantle was provided solely by the rotational constraint based upon the observation of the parameter $\dot{J}_2$ as inferred by Cheng et al. (1989) in the analysis that delivered VM2 from the inversion. Model VM3 was obtained in an inversion that adjusted the observed value of $\dot{J}_2$ so as to allow for the influence of possible contamination due to the influence of modern climate change-induced forcing due to melting from either the Greenland or Antarctic ice sheets. These adjustments were based upon the assumption that one or the other or both of these ice catchments were contributing to the current rate of global sea-level rise at a rate in the range 1–1.5 mm year^{-1}.

the GIA process. **Figure 29** shows Mollweide projections of the global map of this prediction for the ICE-5G and ICE-6G models both including and excluding the impact of rotational feedback. These calculations have been made by solving the SLE using the 'full glacial cycle' methodology for incorporation of the influence of coastline migration discussed in Peltier (2005) and Peltier and Fairbanks (2006). Inspection of the figures for the ICE-5G (VM2) model reinforces the fact that the impact of rotational feedback is such as to superimpose upon the pattern predicted by this model with feedback, a modification that has the form of a spherical harmonic of degree 2 and order 1, a pattern of 'quadrupolar' form. This characteristic of the relative sea-level response to variations in the Earth's rotation was first demonstrated by Dahlen (1976) in the context of his analysis of the 'pole tide' raised in the oceans by the Chandler wobble of the Earth. Initial attempts to incorporate the influence of rotational feedback upon the variations of sea level caused by the GIA process were described by Milne and Mitrovica (1996) and Peltier (1998) who employed the theory of Peltier (1982) and Wu and Peltier (1984) to estimate the magnitude of the effect, analyses that initially suggested it to be small enough to be neglected. The fact that the influence of rotational feedback has this quadrupolar structure is a consequence of the dominance of the contribution due to the influence of TPW, as will be clear on the basis of inspection of eqns [13]. The ω_3 (nontidal acceleration of rotation-related) component of the perturbation to the angular velocity of the planet controls the impact upon sea-level

Table 2 List of the results for the predictions of the Earth's rotation observables for the models analyzed in this paper and the results for the two Stokes coefficients of degree 2 and order 1 in the analyses of GRACE time-dependent gravity data published by Peltier and Luthcke (2009)

Model	*PW speed deg 10^{-6} years*	*PW angle*	$\dot{J}_2 \times 10^{11}$	$\dot{C}_{21} \times 10^{11}$	$\dot{S}_{21} \times 10^{11}$
ICE-6G (VM5aD1)	1.03	−77.69	−3.25	−0.32	1.45
Model A = ICE-6G (VM5aD1) + Grn + Ant + AK	1.64	−57.76	+0.22	−1.24	2.00
Models consisting of model A augmented by melting in SA + Him + Cordillera					
Model B	0.67	−59.63	−1.46	−0.49	0.86
Model C	0.77	−65.36	−1.45	−0.46	1.03
Model D	0.77	−65.97	−1.30	−0.46	1.04
Model E	0.79	−68.43	−1.14	−0.41	1.08
Models differing from B–D only by radial viscosity model (VM5a rather than VM5aD1)					
ICE-6G (VM5a)	0.82	−77.89	−2.78	−0.25	1.16
Model A-VM5a	1.44	−55.15	0.48	−1.17	1.70
Model B-VM5a	0.48	−52.69	−1.10	−0.42	0.57
Model C-VM5a	0.57	−61.40	−1.09	−0.39	0.74
Model D-VM5a	0.57	−62.26	−0.95	−0.39	0.75
Model E-VM5a	0.59	−65.66	−0.80	−0.35	0.80
The aforementioned are to be compared with the following: post-1992 inferences of Roy and Peltier (2011)					
Roy and Peltier (2011)	0.50	−58.0	−0.90	−0.40	+0.60

Model A is model ICE-6G (VM5aD1) of Peltier and Drummond (2010) with modern land-ice melting rates for Greenland, Antarctica, and Alaska fixed to the values inferred from GRACE time-dependent gravity data by Peltier (2009). Models B, C, D, and E are characterized by the addition of rates of land-ice melting to those of model A for sources in the Andes, the Himalayas, and the Sierras. Results are also listed for models B, C, D, and E when viscosity model VM5aD1 is replaced by viscosity model VM5a. See the text for a detailed discussion of these results.

Reproduced from Peltier WR, Drummond R, and Roy K (2012) Comment on "Ocean mass from GRACE and glacial isostatic adjustment" by D.P. Chambers et al. *Journal of Geophysical Research: Solid Earth* 117: B11403, http://dx.doi.org/10.1029/2011JB008967.

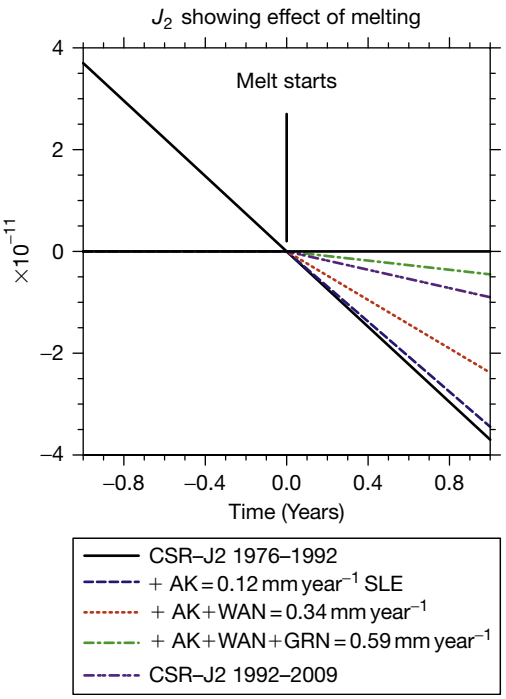


Figure 28 Based upon the assumption that a GIA model has been constructed that represents only ice-age influence and that fits the pre-1992 $\dot{J}_2$ constraint in this figure, this figure illustrates the impact to be expected upon $\dot{J}_2$ of adding the forcing associated with melting of the Greenland, Antarctic, and Alaskan accumulations of land ice as constrained by the GRACE inferences of Peltier (2009). Although this comparison suggests that the change in the secular rate $\dot{J}_2$ could be explicable solely in these terms, it is demonstrated in the succeeding text that this explanation is inadequate.

history of the GIA-induced change in the l.o.d.. The ω_1 and ω_2 perturbations, on the other hand, control the polar wander contribution, and it will be clear on the basis of eqns [13] that this will appear as a forcing of spherical harmonic degree 2 and order 1 form. Noticeable also on the basis of **Figure 29** is the fact that the magnitude of this impact is significantly more intense in the previous ICE-5G (VM2) model than it is in the more recent ICE-6G (VM5a) model, a consequence primarily of the fact that the degree 2 and order 1 Stokes coefficients for this model are larger than those for the ICE-6G (VM5a) model. This amplification of the values of these Stokes coefficients in the earlier model arose as a consequence of a fault associated with an issue fully discussed in Peltier et al. (2012) connected to the need to 'renormalize' the representation of the rotational feedback contribution to these coefficients. The difference in the magnitude of these coefficients for these two models can be exploited to examine the extent to which the magnitude of these Stokes coefficients impacts the fit to relative sea-level observations in certain regions of the Earth's surface. Since it is in the regions centered within the 4 'bull's-eyes' of the spherical harmonic degree 2 and order 1 quadrupole pattern that the influence of rotational feedback will be most apparent, it is in these regions that we should expect to find its consequences most clearly revealed.

The most important region in which such comparison is possible concerns the extremum in rotational feedback that is located on the southern tip of the South American continent (see **Figure 20** shown previously). Although the west coast of this continent is strongly impacted by the deformation associated with the subduction of the Nazca Plate, the east coastal margin is passive and therefore should provide an excellent region in which to attempt to measure, using postglacial histories of relative sea-level change, the impact of rotational

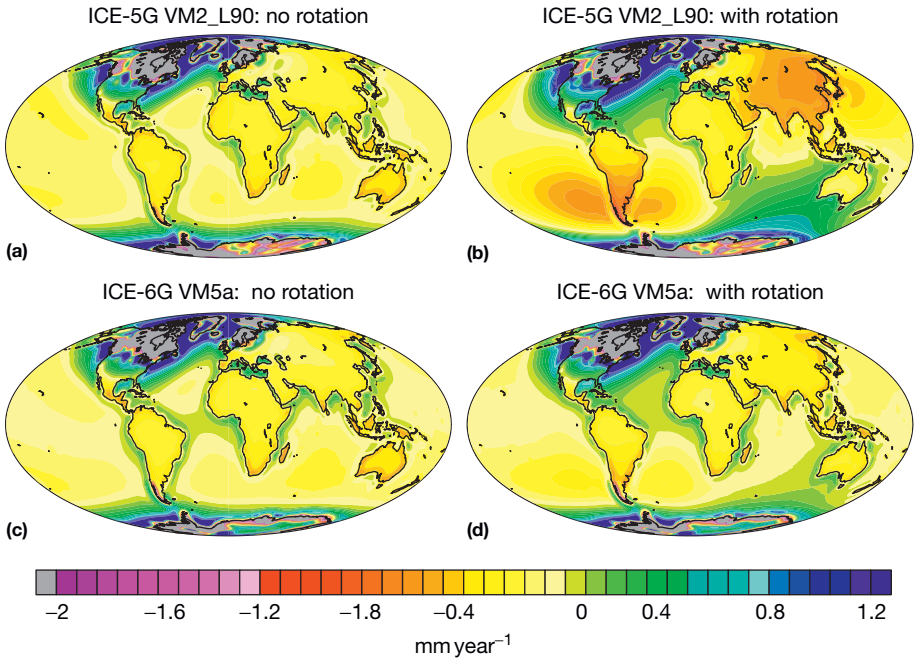


Figure 29 Comparison of the predictions of the ICE-5G (VM2) with those of ICE-6G (VM5a) models of the present-day rate of relative sea-level rise both including and excluding the influence of rotational feedback as described by the second term in the integrand of the triple convolution integral in eqn [10]. Noticeable by inspection of this figure is the fact that the influence of rotational feedback has the form of a spherical harmonic of degree 2 and order 1, which exists as a consequence of the dominant role played in the feedback process by the polar wander component of the rotational response to the GIA process (see eqns [36a] and [36b]). Also evident is the fact that this feedback is stronger in model ICE-5G (VM2) than it is in ICE-6G (VM5a). This is a consequence of the fact that the degree 2 and order 1 Stokes coefficients of the latter model are significantly smaller than those of the former, in fact by a factor of 2.51 as we will see in the succeeding text. Comparing this figure with the equivalent in Peltier (2007a,b) will demonstrate that the results for model ICE-6G (VM5a) are similar to those of the previous model ICE-4G (VM2) as the latter model is characterized by degree 2 and order 1 Stokes coefficients that are very close to those of the ICE-6G (VM5a) model.

feedback. Figure 30 shows a map of the South American continent, from Peltier (2005), on which, primarily, along the east coast passive continental margin, the locations are shown at which, based upon the compilation of relative sea-level histories in Rostami et al. (2000), radiocarbon-dated relative sea-level history information is available. When such data are used to compare with the predictions of the theory previously described, they must of course be transformed onto the calendar year timescale using the modern calibration procedure CALIB described by Stuiver and Reimer(e.g., 1993 and subsequent releases).

In Rostami et al. (2000), attention was drawn to the fact that there seemed to exist a systematic misfit of the theory for postglacial sea-level change based upon the ICE-4G (VM2) model without rotational feedback to the observed sea-level histories from this coastal region. The nature of this misfit, documented in Figure 11 of Rostami et al. (2000), involved the variation with latitude of the so-called mid-Holocene 'high stand' of the sea. The prediction of the amplitude of this high stand using the ICE-4G (VM2) model without feedback was that this amplitude should decrease with increasing south latitude, whereas the observation was that it actually increased very significantly, a qualitative difference easily resolved by the observational record if the data were being interpreted correctly. In Peltier (2002c), it was first demonstrated that the

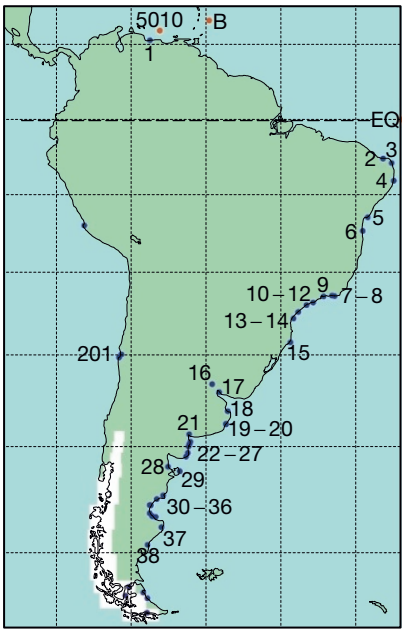


Figure 30 Location map for the South American sites from which radiocarbon-dated relative sea-level histories are available. Also shown in the figure is the LGM location of the Patagonian ice-sheet complex.

incorporation of rotational feedback into the theory would lead to the qualitative change in the nature of the prediction required to reconcile the observations, but significant discrepancies remained for the data from the southernmost section of the coast. More recently, Peltier (2005) has more fully exploited these observations and the ICE-4G (VM2) model predictions to argue that this data set could be invoked to demonstrate that the impact of rotational feedback upon RSL history predicted by the ICE-4G (VM2) model was approximately correct. In fact, the data–model comparisons presented in both Peltier (2002c) and Peltier (2005) were incomplete, as several of the most anomalous of the observations could not be reconciled.

The importance of including the ICE-5G calculations together with those for the ICE-6G model for present purposes is that the former model is characterized by an excessively large contribution from the degree 2 and order 1 Stokes coefficients as previously mentioned. This model may therefore be seen as a surrogate for models that share this characteristic. By comparing the predictions to South American east coast observations, we will be able to isolate the extent to which, as well as the sense in which, the magnitude of the degree 2 and order 1 Stokes coefficients directly impacts sea-level history in this critical region. Although we will not discuss this additional sensitivity here, the influence of a modest deviation from the 'equivalent Earth model' assumption of Munk and MacDonald (see Peltier and Luthcke (2009) for further discussion) is also of interest although, as explicitly demonstrated in Peltier and Luthcke (2009), this influence appears to be negligible in comparison with the more important influences under discussion.

Figure 31 presents model–data comparisons for postglacial relative sea-level history at 16 of the numbered sites from the east coast of South America, beginning at the northernmost site from coastal Venezuela and continuing southward through Brazil and Argentinean Patagonia towards Tierra del Fuego. For each site, the predictions for three different models are shown. These are for the ICE-5G (VM2) model that includes the influence of rotational feedback and for the ICE-6G (VM5a) model both including and excluding this influence. It will be important to note that the first site in the list, denoted NC Venezuela (Valastro et al., 1980), is the only site that is north of the equator (see **Figure 30**). This is important because the sign of the impact of rotational feedback changes across

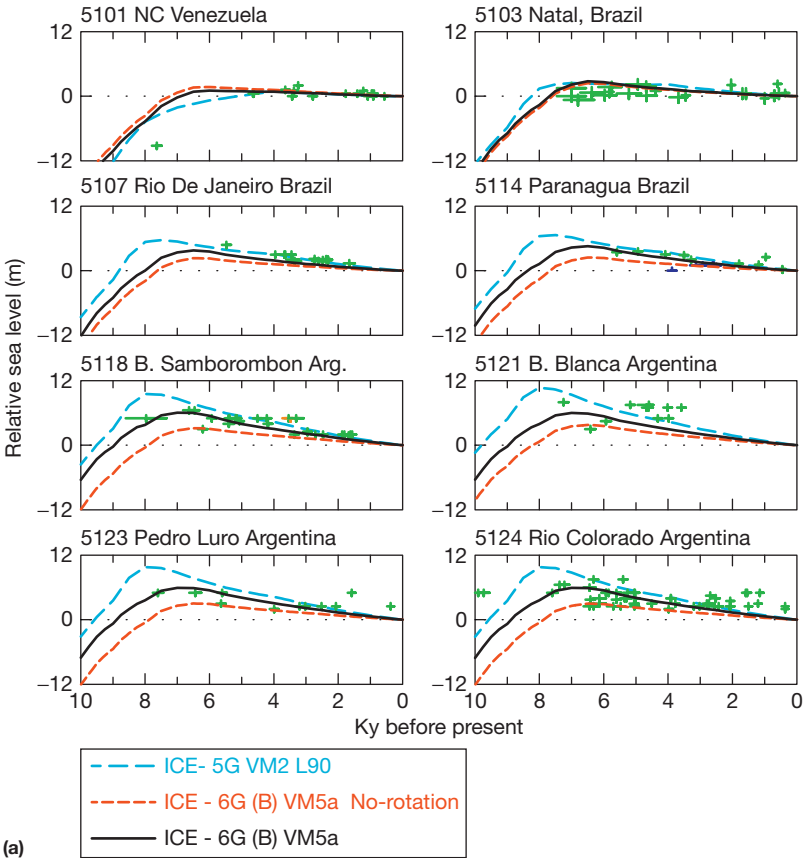


Figure 31 (a) Model–data intercomparisons for eight sites from the northern portion of the east coast of the South American continent. The last two digits of the site numbers correspond to the numbered locations in **Figure 30**. On each plate, the ^{14}C-dated sea-level index points are denoted by green crosses. RSL history predictions based upon eqn [10] are shown for three model variants, respectively, ICE-6G (VM5a) with rotational feedback (black), ICE-6G (VM5a) without rotational feedback (red), and ICE-5G (VM2) with rotational feedback (mauve). Notable based upon the comparisons at these north coastal sites is that the previous ICE-5G (VM2) model with rotational feedback has an excessive amplitude of this influence, whereas the ICE-6G (VM5a) model has a more modest rotational feedback effect, which is in much closer accord with the data. In the absence of rotational feedback, however, one cannot fit the observations of the mid-Holocene high stands of the sea that are recorded in this region.

(Continued)

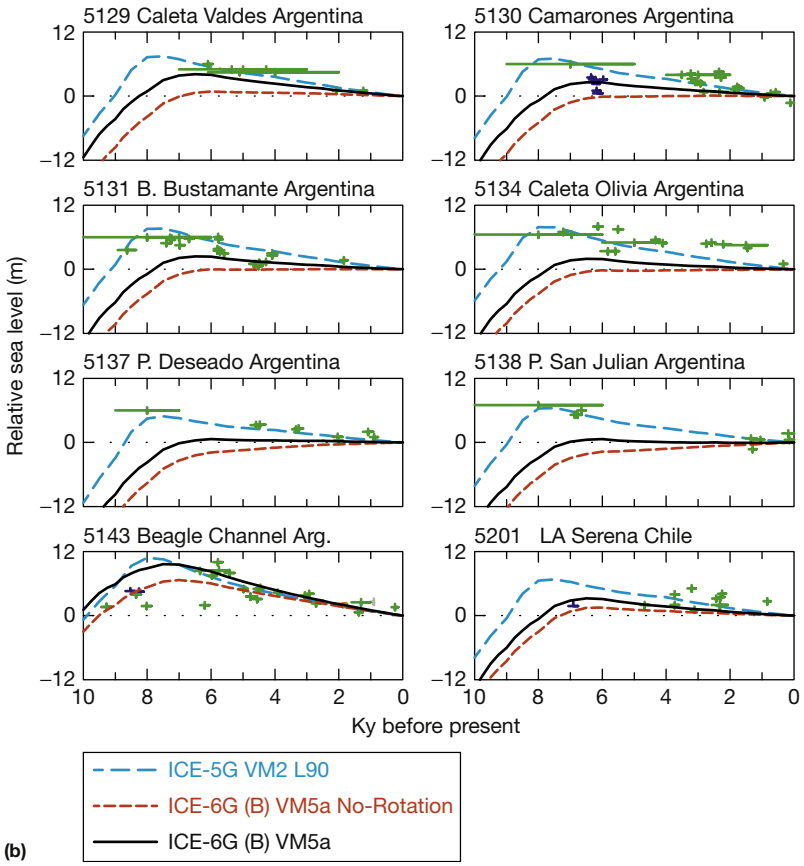


Figure 31, Cont'd (b) Same as (a) but for a set of eight sites from the southern (Patagonian) region of the east coast of the South American continent. At these locations, the results of the intercomparisons strongly conflict with those from the northern segment of the coast in strongly preferring the very large-amplitude rotational feedback effect over the weaker influence due to this cause that is embodied in the ICE-6G (VM5a) model and the precursor model ICE-4G (VM2).

this line of 0° latitude, as expected based upon the spherical harmonic degree 2 and order 1 form of the pattern of this influence. To the north of the equator, as will be clear by inspection of the comparisons at NC Venezuela, the impact of rotational feedback is to diminish the predicted elevation with respect to sea level throughout the Holocene interval from 10 ka onward. To the south of the equator, however, as the comparisons for the Natal, Brazil, site (Bezerra et al., 2003) show, the opposite is the case, as at this site, the impact of rotational feedback is such as to elevate the predicted sea levels for the Holocene epoch above that for which this influence is neglected entirely. Based upon the difference in the values of the degree 2 and order 1 Stokes coefficients for the ICE-5G and ICE-6G models as will be discussed in detail in the following section of this chapter, it will be clear that the influence of rotational feedback is expected to be weaker in ICE-6G (VM5a) than it is in ICE-5G (VM2). As will be evident by inspection of the complete set of comparisons shown in Figure 31, this hierarchy is preserved at all locations. Table 3 lists the references for the data from all of the sites for which data–model comparisons are being shown in this section.

Very clearly evident also, based upon the results shown in Figure 31, is the fact that exceptionally high 'stands' of the sea above present sea level are predicted and observed to obtain at sites along this coast during mid-Holocene time between 6000 and 8000 years before present. Although there is considerable scatter in the data at the sites from NC Venezuela to Rio Colorado, Argentina, for which comparisons are shown in Figure 31(a), it is nevertheless clear that the influence of rotational feedback is required in order for the model to adequately reconcile the data. In the absence of this influence, the amplitude of the recorded mid-Holocene high stands is insufficient. As discussed in detail in Rostami et al. (2000), most of the data from this set of locations are from the older literature, and therefore, a legitimate question exists concerning its quality. A general conclusion that nevertheless follows for the comparisons in Figure 31(a) is that the predictions of the previous ICE-5G (VM2) model of the amplitude of the mid-Holocene high stands at these locations are excessive. This is encouraging because we expect that the degree 2 and order 1 Stokes coefficients for the ICE-6G (VM5a) model are more accurate than those for the previous ICE-5G (VM2) model for the Holocene interval.

Considering next the comparisons for the collection of sites shown in Figure 31(b), the situation is reversed, in that it is only the very large-amplitude high stand predictions of the ICE-5G (VM2) model that are able to reconcile the data from these more southerly locations. Even for the more recently collected data from this region, however, significant issues are evident. Concerning the data from the set of locations between

Table 3 References for the data from the sites for which data–model intercomparisons are provided in this chapter

Site number	*Site name*	*References*
5101	NC Venezuela	Valastro et al. (1980)
5103	Natal, Brazil	Bezerra et al. (2003)
5107	Rio de Janeiro, Brazil	Martin et al. (1987), Fairbridge (1976), Delibrias et al. (1974), and Angulo and Lessa (1997)
5114	Paranaguá, Brazil	Angulo and Lessa (1997)
5118	B. Samborombón, Argentina	Codignotto et al. (1992)
5121	B. Blanca, Argentina	Codignotto et al. (1992)
5123	Pedro Luro, Argentina	Albero and Angiolini (1983, 1985)
5124	Rio Colorado, Argentina	Codignotto et al. (1992)
5129	Caleta Valdés, Argentina	Rostami et al. (2000), Codignotto et al. (1992), and Albero and Angiolini (1983)
5130	Camarones, Argentina	Rostami et al. (2000)
5131	B. Bustamante, Argentina	Rostami et al. (2000)
5134	Caleta Olivia, Argentina	Rostami et al. (2000)
5137	P. Deseado, Argentina	Rostami et al. (2000)
5138	P. San Julián, Argentina	Rostami et al. (2000)
5143	Beagle Channel, Argentina	Porter et al. (1984) and Morner (1991)
11105	Pt. Adelaide/ Gillman, Australia	Belperio et al. (2002)
11107	Pt. Wakefield, Australia	Belperio et al. (2002)
11109	Pt. Pirie (Spencer Gulf), Australia	Belperio et al. (2002)
11110	Redcliffe, Australia	Belperio et al. (2002)
11111	Franklin HB, Australia	Belperio et al. (2002)
11113	Ceduna, Australia	Belperio et al. (2002)
11115	Weiti River, New Zealand	Gibb (1986) and Schofield (1975)
111123	Blue Skin Bay, New Zealand	Gibb (1986)
9003	Nouakchott, Mauritania	Faure and Hebrard (1977)
9006	St. Louis, Senegal	Faure et al. (1980)
9022	Reunion	Camoin et al. (1977, 2004)
9023	Mauritius	Camoin et al. (1997)
11004	Shikoku island, Japan	Pirazzoli (1978)
11006	Kikaijima island, Japan	Pirazzoli (1978) and Sugihara et al. (2003)
11008	S. Okinawa island, Japan	Koba et al. (1982)
11010	Uotsuri island, Japan	Koba et al. (1982)

Caleta Valdes, Argentina, and P. San Julian, Argentina, all from Argentinean Patagonia, it will be evident that at several of these locations, only a single data point exits rather than a continuous relative sea-level history. At these locations, although the data are thought to have been carefully screened for quality (see discussion in Rostami et al. (2000)), it is clear on the basis of them that only the ICE-5G (VM2) model is able to reconcile the 6–8 m mid-Holocene high stands of the sea that are observed to exist in the time window between 6 and 8 thousand years before present. The ICE-6G (VM2) model, in the absence of rotational feedback, inevitably predicts that no such high stand should exist or, with feedback, that its amplitude is unacceptably small. The ICE-5G (VM2) model predicts a high stand amplitude that does fit these extremely high isolated RSL data points, but this model is known to be characterized by values of the degree 2 and order 1 Stokes coefficients that are now understood to be exaggerated by a factor of precisely 2.51 for viscosity model VM2 (see the detailed discussion in Peltier et al. (2012)). These analyses very strongly reinforce the conclusion in Peltier (2005, 2007) that the influence of rotational feedback is apparently crucial to understanding the relative sea-level history data from along this coast. However, neither the VM2 nor the VM5a viscosity model is characterized by Stokes coefficients of degree 2 and order 1 that are large enough to explain the data from the southernmost region of the coast. It remains an open question whether a further variation on the viscosity model or loading history will enable the correction of this characteristic misfit. We are also unable to rule out the possibility that the extremely large-amplitude isolated high stands inferred to exist at the Patagonian sites could correspond to storm beach deposits or to exist as a consequence of the influence of a marked changes in the (palaeo)tidal regime in the range of time during which the extreme changes in relative sea level were recorded (e.g., see Griffiths and Peltier (2009); Hill et al. (2011); and Hall et al. (2013); the latter papers provide analyses for the US east coast that take changes in tidal range into account in the interpretation of radiocarbon-dated RSL histories along this coast).

It is nevertheless the case that there is scatter in the data, and one might reasonably hope that there might exist data from other locations where the influence of rotational feedback is also expected to be strong that might be invoked to confirm (or to deny) the robustness of the conclusion that the influence of rotational feedback upon the GIA process is significant and is required to understand the available observational constraints. To this end, **Figure 32** shows a location map for 16 additional sites from other regions, for some of which, this influence is expected to be similarly evident. Eight of these sites are from the New Zealand–Australia region, for which comparisons with model predictions are shown in **Figure 33**. These sites are within the influence of the second southern hemisphere 'bull's-eye' of the spherical harmonic degree 2 and order 1 pattern that is produced by the influence of rotational feedback. The map in **Figure 32(c)** (in Mercator projection), upon which all of these additional positions are located, is that of the difference between the present-day rate of sea-level rise prediction for the ICE-6G (VM5a) model with rotation (shown previously in **Figure 32(a)**) and that obtained when this influence

is not included (shown in Figure 32(b)). The eight additional comparisons shown in Figure 34 are for sites of which 4 are from locations in the Indian Ocean or the west coast of Africa and 4 of which are from the third 'bull's-eye' located over the Japanese islands (see Figure 32(c)).

Inspecting first the comparisons shown in Figure 33 for sites in the Australia–New Zealand region will show that at all of these sites, there exists a striking anomaly between the predictions of the version of the ICE-6G (VM5a) model without rotational feedback and the predictions for either of the models in which this influence is included. In the absence of rotational feedback, the theory predicts that a mid-Holocene high stand of the sea should exist with an ~3 m amplitude. However, the data at all 8 locations show that no such feature is evident. At most of the locations for which results are shown in Figure 33, either of the two models with rotational feedback included significantly reduces the amplitude of this predicted feature (ICE-6G (VM5a)) or eliminates it entirely (ICE-5G (VM2)). Again, only the latter model with the excessive values of the degree 2 and order 1 Stokes coefficients fits the observations as was the case along the southeast coast of the South American continent. This set of comparisons is especially important because in this region, the sign of the feedback effect is predicted to be opposite to that operating over the South American continent. Nevertheless, the necessity of the operation of a strong influence of rotational feedback in order to fit the observational constraints remains clear.

The further model–data comparisons shown in Figure 34 include information from sites in which the sign of the influence of rotational feedback is the same as that that operates in the Australia–New Zealand region and sites in which the influence is qualitatively similar to that evident at locations along the southeast coast of the South American continent. Sites of the former type include those along the west coast of Africa (Nouakchott, Mauritania, and St. Louis, Senegal) and the others from the Indian Ocean (Réunion and Mauritius). Sites of the latter type from the islands of Japan include Shikoku island, Kikaijima island, S. Okamura island, and Uotsuri island. At the former locations, theory once more predicts that in the absence of rotational feedback, a mid-Holocene high stand of the sea should be observable. At these locations, however, no such high stand is evident. The addition of rotational feedback, of the excessively high strength embodied in ICE-5G (VM2), once more allows the model to fully reconcile the data, whereas although the predictions of the follow-on ICE-6G (VM5a) model demonstrate that there should be a nonnegligible impact of rotational feedback at these locations, the strength of the feedback appears to be weaker than required.

Turning next to the final set of four data–model comparisons shown in Figure 34, it will be clear that at each of the Japanese islands for which relative sea-level comparisons are shown, the model without rotational feedback predicts that no mid-Holocene high stand of the sea should be evident at any of these locations. Yet, the data in each case clearly reveal the presence of such a feature. As will be clear by inspection of this set of comparisons, the incorporation of the influence of rotational feedback suffices to reconcile the misfits of the theoretical predictions to the data that would otherwise exist for the ICE-5G (VM2) model in which this feedback is excessively

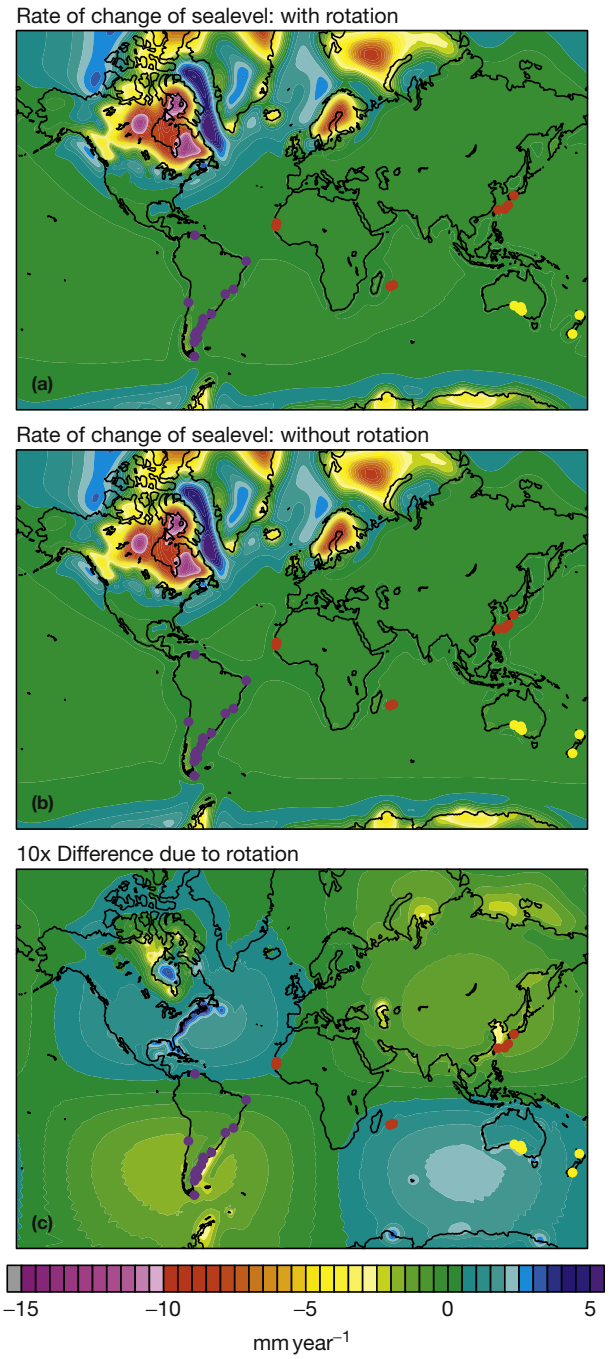


Figure 32 Mercator projections of the model-predicted present-day rate of relative sea-level rise for the ICE-5G (VM2) model both including (a) and excluding (b) the influence of rotational feedback. In (c), the difference between (a) and (b) is presented, demonstrating once more that the influence of the rotational feedback effect is dominantly of the form of a spherical harmonic of degree 2 and order 1. Superimposed upon each of the three plates of this figure are the locations of sites at which high-quality ^{14}C-dated RSL histories are available beyond those from the South American continent already discussed. The sites shown as yellow dots are located in a region (Australia–New Zealand) where the influence of rotational feedback is expected to have the opposite sign as the South American locations. At the sites shown as red dots, the effect is either expected to the same as in Australia–New Zealand (in the Indian Ocean or West Africa) or the same as in South America (Japanese islands).

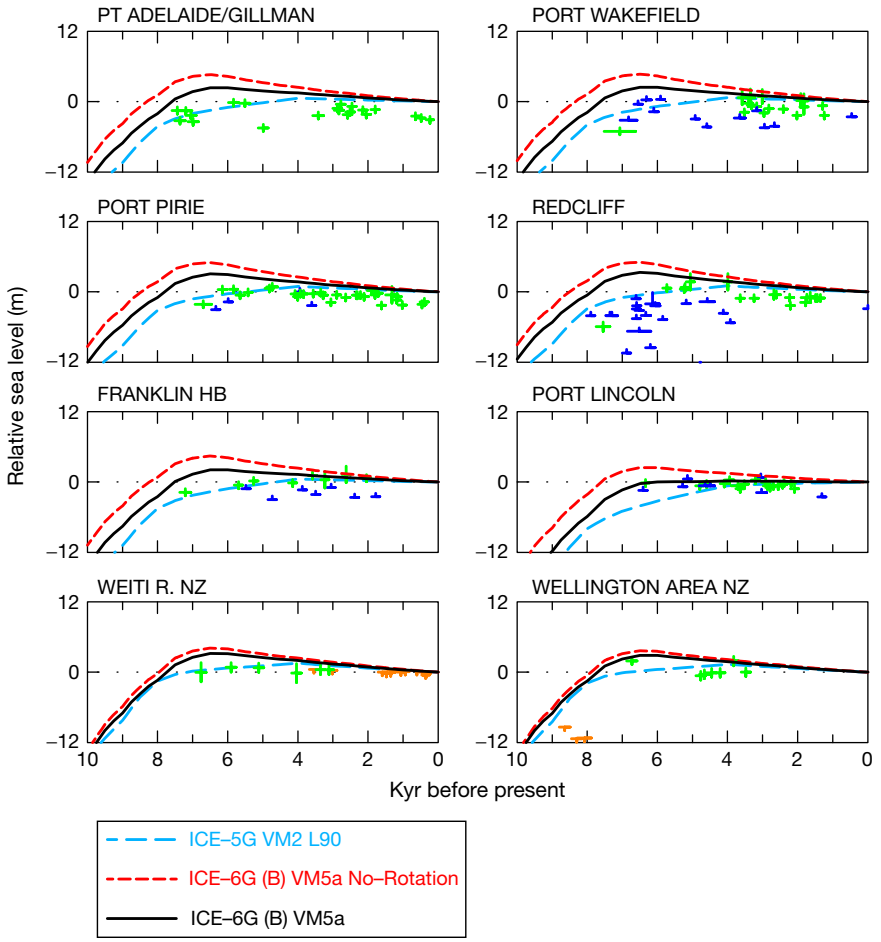


Figure 33 Same as **Figure 31(a)** but showing model–data comparisons for the Australia–New Zealand region. Inspection of these comparisons demonstrates that, as expected, the nature of the influence of rotational feedback is opposite to that observed at the South American locations. At these locations, the data once more appear to strongly prefer the model with the overly strong rotational feedback embodied in ICE-5G (VM2).

strong, but even in the ICE-6G (VM5a) model for which the feedback is weaker, it remains a nonnegligible influence.

On the basis of the totality of these comparisons, we may therefore conclude that there exists very strong evidence in the global record of postglacial sea-level history of the action of the influence of rotational feedback. However, the strength of this feedback in the new ICE-6G (VM5a) model of the GIA process appears to be somewhat weak, as it was in the precursor model ICE-4G (VM2). In fact, this is not surprising as the procedure employed to construct the ICE-4G (VM2) model did not suffer from the same technical flaw that arose in the course of the ICE-5G (VM2) reconstruction process. In the ICE-5G (VM2) model, however, in which the larger magnitude of the critical Stokes coefficients of degree 2 and order 1 was excessive for the VM2 viscosity profile, the strength of the required feedback appears to be just what is required to best fit the observations. It remains to be seen whether it will be possible to further refine the radial profile of mantle viscosity or the ice-sheet loading history in such a way as to increase the values of the Stokes coefficients for ICE-6G (VM5a) without violating additional constraints. That this could be possible follows from the fact that the rotational anomalies depend only upon the viscosity of the deepest mantle not upon the shallower structure to which relative sea-level histories are uniquely sensitive.

9.09.6 Grace Satellite Inferences of Geoid Height Time Dependence: The Combined Influences of Ice Age and Modern Land-Ice Melting

At several points in the previous section of this chapter, I have commented on the importance of the time-dependent geoid as a representation of the strength and spatial characteristics of the ongoing process of global GIA. With the launch of the GRACE satellites in March of 2002, it became possible to directly measure an approximation to this field globally and to begin to disentangle the contributions to it from ice-age-related GIA and from the modern climate change-related processes that also contribute. In this section of this chapter, my purpose is to review a series of recent results that have contributed to refining our understanding of geoid height time dependence and the approximation to it that the GRACE system is

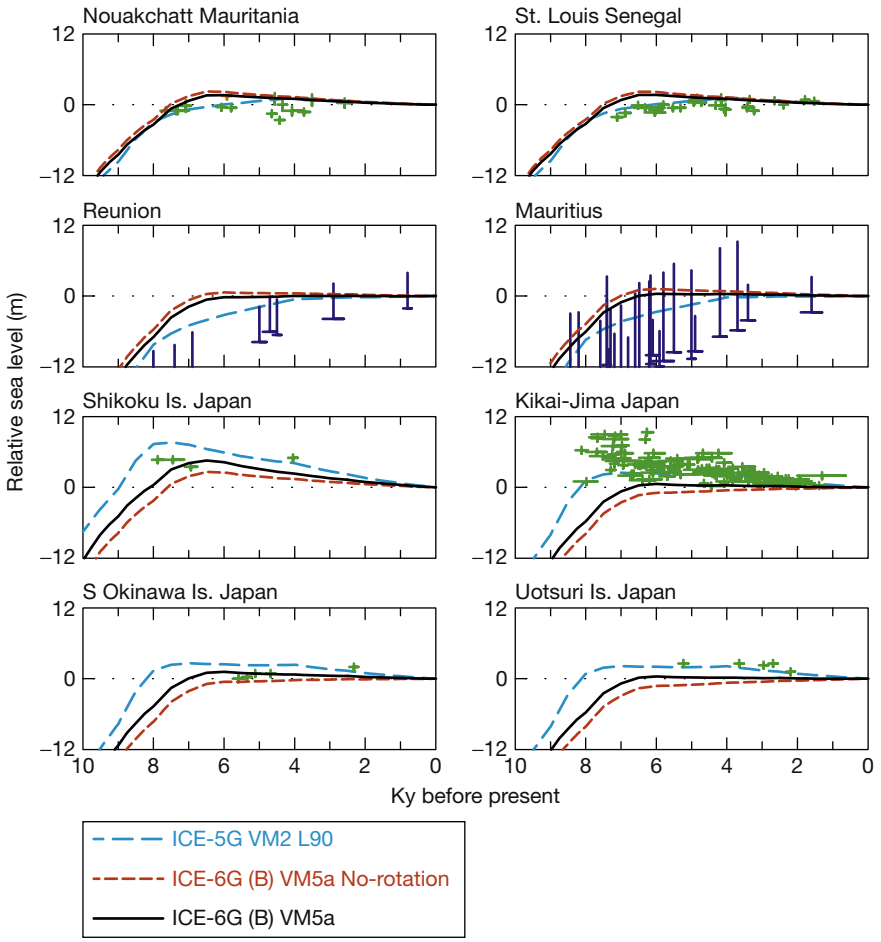


Figure 34 Same as for **Figure 31(a)** but showing model–data comparisons for both 'down sites' (Indian Ocean and West Africa) and 'up sites' (Japanese islands). For all sites, the sign of the rotational feedback effect is as expected and most data appear to prefer model ICE-5G (VM2), which embodies excessively strong rotational feedback.

actually able to observe, as these results are proving to be critical to our ability to make full use of the measurements, especially insofar as the extraction of modern climate change-related information is concerned.

To begin, it is important to be reminded of the formal definition of the geoid of classical geodesy as the surface of constant gravitational plus centrifugal potential that overlaps the surface of the sea in the absence of currents and tides. When we solve the sea-level equation [10] to obtain the history of relative sea-level change to be expected anywhere on the Earth's surface due to the time-dependent exchange of mass between the oceans and land-based ice sheets, the solution is explicitly constructed so as to ensure that the new position-dependent and time-dependent sea level at every instant of time lies on such a globally defined geoid. Although the gravitational plus centrifugal potential characteristic of this geoid varies with time as mass is exchanged between the oceans and the continents and as the rotational state of the planet evolves, the ocean surface by construction remains such an equipotential. Now, this representation through the SLE of the time-evolving geoid is an approximation for a number of reasons. First, it is based upon the assumption that the system is evolving sufficiently slowly that the quasistatic approximation that lies at the heart of the SLE is an adequate approximation. This neglects the dynamic processes that are actually responsible for redistributing the water among the ocean basins that is required to produce a sea surface that is in fact equipotential. Furthermore, of course, there will exist a steric impact upon sea level that also acts to raise or lower the level of the sea during the climate changes that are responsible for inducing the exchange of mass between the oceans and the continents. When we employ the solutions of the SLE based upon eqn [10] to predict sea-level history, we are assuming that the steric overprint upon the mass effect is sufficiently small as to be negligible. This issue has never been adequately explored but can be using an appropriately modified version of the SLE formalism coupled to a full glacial cycle integration of a modern coupled atmosphere–ocean climate model. Since current computational resources remain inadequate to the latter task, it will be some time before this might reasonably be attempted.

9.09.6.1 The Geoid Equation: Analytic Results for the Stokes Coefficients of Degree 2 and Order 1

Since the classical geoid is defined in terms of sea level, it should be clear that, notwithstanding the approximations

upon which the SLE is based, it must provide the basis for a prediction of its time dependence. If the geoid is to be observed from space, it is also clear that the sea level that it represents will be measured not with respect to the surface of the solid Earth but rather with respect to the planet's center of mass relative to which the orbit of an Earth satellite is represented. This shift of the datum with respect to which the evolving sea level is measured is easily accommodated in the SLE formalism. Using this apparatus, we may define the space–time evolution of the quasistatically evolving geoid as

$$G(\theta,\lambda,t)=\left[\int_{-\infty}^{t}\mathrm{d}\tau\iint_{\Omega}\mathrm{d}\Omega'\left\{L(\theta',\lambda',\tau)G_G^L(\phi,t-\tau)+\Psi^R(\theta',\lambda',\tau)_RG_G^T(\phi,t-\tau)\right\}+\left\{\frac{\Delta\Phi(t)}{g}\right\}\right] \quad [45]$$

In this equation, the geoid Green function for the surface load is

$$G_G^L(\phi,t)=\frac{a}{M_e}\sum_{l=0}^{\infty}\left(\delta(t)+k_l^L(t)\right)P_l(\cos\phi) \quad [46]$$

in which 'a' is the Earth's radius and M_e is the Earth's mass. The corresponding renormalized Green function for the impact upon the geoid associated with the changing rotation is, correspondingly,

$${}_RG_G^T(\phi,t)=\frac{1}{g}\sum_{l=0}^{\infty}\left(\delta(t)+k_l^T(t)\right)\frac{2l+1}{4\pi}P_l(\cos\phi) \quad [47]$$

in which 'g' is the acceleration of gravity at the Earth's surface and the factor $(2l+1)/4\pi$ is the renormalization factor previously discussed in connection with eqn [10]. The parameters $k_l^L(t)$ and $k_l^T(t)$ in eqns [40] and [47], as before, are the viscoelastic surface mass load and tidal potential load Love numbers, which enable the aforementioned Green functions, when convolved with the surface mass load and tidal potential load, respectively, to translate these applied 'forcings' into the impacts upon the surface gravitational plus centrifugal potential field (the geoid) due to both the instantaneous effect of the forcing and the induced internal redistributions of mass that it causes. Insofar as the geoid is concerned, however, these variations of potential are now measured with respect to the center of mass of the planet rather than with respect to the surface of the solid Earth. The only difference between a solution for relative sea-level history from eqn [10] and a solution for geoidal history from eqn [39] is therefore a shift in the local datum from the surface of the solid Earth in the former case to the center of mass in the latter case. The Green functions that appear in eqns [14] differ from those of eqns [46] and [47] only in that their expansions include additional contributions to the degree-dependent amplitudes that involve the negative of the appropriate Love number for radial displacement (see the succeeding text). The theory required to construct the time-dependent Love numbers was first articulated in Peltier (1974) and Peltier and Andrews (1976). In the Green function expression for the surface load, the addition theorem for spherical harmonics (e.g., Mathews and Walker, 1987) may be invoked to write

$$P_l(\cos\gamma)=\frac{1}{(2l+1)}\sum_{m=-l}^{+l}Y_l^{m\,*}(\Omega')\,Y_l^m(\Omega) \quad [48]$$

The issue as to whether the space-independent but time-dependent conservation of mass factor, $\Delta\Phi(t)/g$, which is clearly required in the SLE, should also appear in the geoid eqn [45] is an issue that was raised in Chambers et al. (2010) in connection with the use of the equation to compare GIA predictions of the time-dependent geoid with the GRACE satellite observations, and so, this term has been placed in brackets in eqn [45]. The term is of course required in the description of the true sea level-related geoid as discussed in detail in Peltier et al. (2012).

Although the space–time-dependent geoid may be computed directly from eqn [45], it may also be obtained by first solving the sea-level equation [10] to obtain the spherical harmonic expansion of the time-dependent relative sea-level history $S(\theta,\lambda,t)$. At every point on the surface, one may then compute its time derivative, $\dot{S}(\theta,\lambda,t)$ say. The time dependence of geoid height may then be obtained by next computing the radial displacement of the solid Earth, $R(\theta,\lambda,t)$ say. The latter calculation is performed by evaluating the triple convolution integral:

$$R(\theta,\lambda,t)=\int_{-\infty}^{t}\mathrm{d}t'\iint_{\Omega}\mathrm{d}\Omega'\left\{L(\theta',\lambda'',t')G_R^L(\phi,t-t')+\Psi^R(\theta',\lambda',t')_RG_R^T(\phi,t-t')\right\} \quad [49]$$

in which the Green functions for the radial displacement due to the surface mass load and due to the variation in the centrifugal potential are (Peltier, 1974; Peltier and Andrews, 1976)

$$G_R^L(\phi,t))=\frac{a}{M_e}\sum_{l=0}^{\infty}h_l^L(t)P_l(\cos\phi)$$

$${}_RG_R^T(\phi,t)=\frac{1}{g}\sum_{l=0}^{\infty}h_l^T(t)\frac{2l+1}{4\pi}P_l(\cos\phi) \quad [50a,b]$$

in which $h_l^L(t)$ and $h_l^T(t)$ are the surface mass load and tidal potential load Love numbers for radial displacement. Given $R(\theta,\lambda,t)$ from eqn [43], we may then compute its time derivative, $\dot{R}(\theta,\lambda,t)$ say. To determine the time dependence of geoid height, we may then simply compute, as originally discussed in Peltier (1999) (notwithstanding the issue connected with the time derivative of the mass conservation term in the SLE mentioned previously),

$$\frac{\mathrm{dGeoid}}{\mathrm{d}t}=\dot{S}(\theta,\lambda,t)+\dot{R}(\theta,\lambda,t) \quad [51]$$

Since both 'S' and 'R' and their time derivatives are known in terms of spherical harmonic expansions, eqn [51] also provides a spherical harmonic expansion for geoid height time dependence. In all previous calculations of this field by the Toronto group, the degree 2 and order 1 Stokes coefficients have been determined together with all other spherical harmonic coefficients directly from eqn [51]. The reader may well wonder why this two-step procedure was employed rather than working with eqn [45] for the geoid directly. The reason for this is simply that 'S' and 'R' and their time derivatives are all required for comparison with various observational

constraints. For example, 'S' is itself a relative sea-level history such as those discussed in the previous section of this chapter that may be compared to an observed history at any position on Earth's surface from which data are available, whereas $\dot{S}$ for the present day is a rate of relative sea-level rise that may be employed to correct a modern tide gauge measured rate for the influence of the GIA process. Similarly, $\dot{R}$ may be employed to compare with a GPS-measured rate of radial displacement of the Earth's solid surface, whereas R defines the instantaneous variation in the radius of the solid Earth (i.e., the time-dependent depression or uplift of the crust).

Figure 35 shows examples of the global fields $\dot{S}$, $\dot{R}$, and dGeoid/dt for the version of the same ICE-6G (VM5aD1) model employed to construct one of the polar wander paths shown in Figure 11. Clearly evident by inspection of Figure 35 is the fact that the field of geoid height time dependence contains a large-amplitude degree 2 and order 1 component due to the influence of rotational feedback, which has been discussed at length in previous sections of this chapter. One of the four extrema of this pattern is centered upon the east coast of the continental United States as is also evident in previous graphics including Figure 32. The second is that centered on the southern tip of the South American continent mentioned previously in connection with our discussion of South American east coast sea-level records. In the recent papers by Horton et al. (2009) and Engelhart et al. (2010, 2011), analyses have focused upon the extent to which relative sea-level data from the US east coast were impacted by the rotational feedback process, adding additional insight to that provided by the discussion in the previous section.

9.09.6.2 The Exact Geoid and the Approximation to It Observable by GRACE

Our further interest in this subsection is to provide exact analytic expressions for the degree 2 and order 1 Stokes coefficients that govern the degree of dominance of the degree 2 and order 1 overprint on the geoid height time dependence evident in Figure 35. The analysis to follow is original to Peltier et al. (2012) (see the supplementary material that accompanied that paper). To begin, we express the surface mass load in a spherical harmonic expansion in terms of a surface mass load

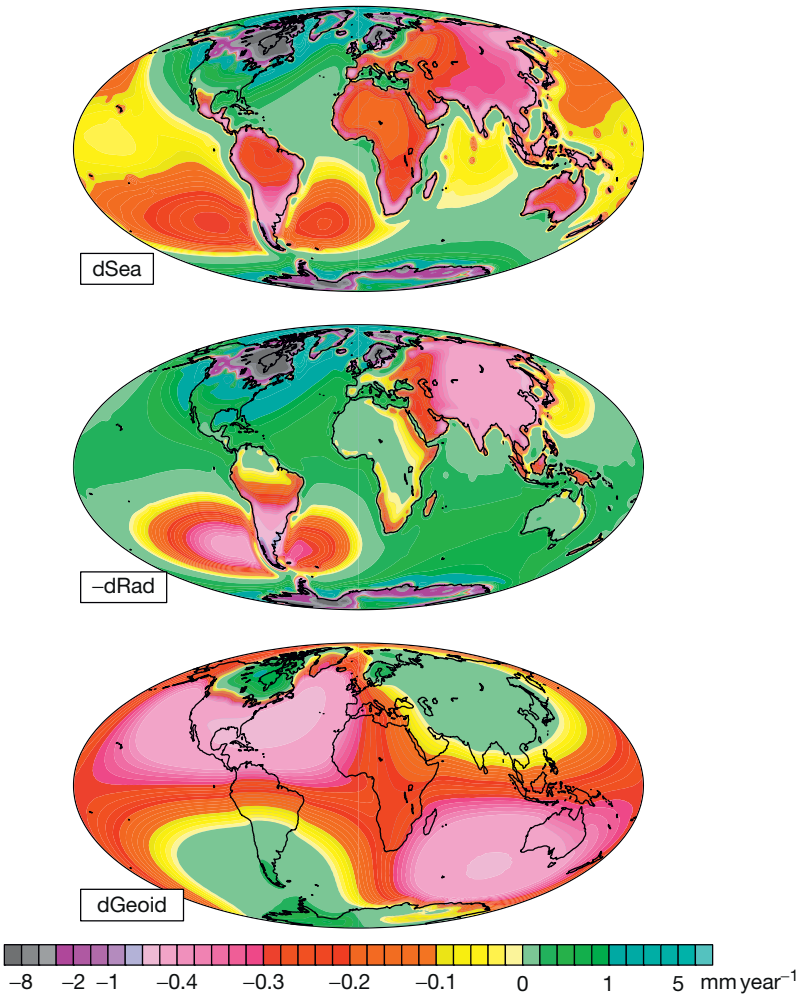


Figure 35 Model ICE-6G (VM5a) results for the present-day rate of relative sea-level rise (dSea), the present-day rate of radial displacement of the Earth's crust (dRad), and the present-day rate of change of geoid height (dGeoid). Notable is the fact that dRad is negatively correlated with dSea. dGeoid is simply the sum of dSea and dRad.

density $\nu(\theta,\lambda,t)$, which includes both the ice component and the bathymetry component such that

$$L(\theta,\lambda,t)=a^2\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\nu_{lm}(t)Y_l^m(\Omega) \qquad [52]$$

in which the spherical harmonics are as in Peltier et al. (2012). If the Earth were entirely rigid, this time-dependent surface mass load would correspond to variations of the elements of the moment of inertia tensor as follows:

$$\begin{aligned}I_{ij}^{\text{Rigid}}(t)&=\int\nu(\theta,\lambda,t)\left(a^2\delta_{ij}-x_ix_j\right)a^2\mathrm{d}\Omega,\ \nu(\theta,\lambda,t)\\&=\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\nu_{lm}(t)Y_l^m(\theta,\lambda)\\&\quad x_x=a\sin\theta\cos\lambda,\ x_y=a\sin\theta\sin\lambda,\ x_z=a\cos\theta,\\&\quad \mathrm{d}\Omega=\sin\theta\mathrm{d}\theta\mathrm{d}\lambda\end{aligned} \qquad [53a]$$

So that

$$\begin{aligned}I_{xz}^{\text{Rigid}}-iI_{yz}^{\text{Rigid}}&=-a^4\iint_{\Omega}\nu(\theta,\lambda,t)\sin\theta\cos\theta e^{-i\lambda}\mathrm{d}\Omega\\&=-a^4\iint_{\Omega}\nu(\theta,\lambda,t)\left[\sqrt{2/15}\overline{P}_2^1(\cos\theta)\right]e^{-i\lambda}\mathrm{d}\Omega\\&=-\sqrt{2/15}a^4\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\nu_{nm}(t)\iint_{\Omega}Y_l^m(\Omega)Y_2^{1*}(\Omega)\mathrm{d}\Omega\\&=-4\pi a^4\sqrt{2/15}\sigma_{21}\end{aligned} \qquad [53b]$$

And thus,

$$4\pi a^4\nu_{21}(t)=-\sqrt{15/2}\left[I_{xz}^{\text{Rigid}}-iI_{yz}^{\text{Rigid}}\right] \qquad [53c]$$

For the total inertia components, we have, as discussed in detail previously, $I_{xz}^{\text{Tot}}=I_{xz}^{\text{Load}}+I_{xz}^{\text{Rot}}$. A similar expression holds for I_{yz}^{Tot}. We may rewrite the geoid equation [39] in two distinct ways, either as a sum of parts that derive from a spherical harmonic expansion in terms of complex spherical harmonics or in the form of an expansion in terms of real spherical harmonics, the latter form being that in which the conventional Stokes coefficients appear directly, as follows:

$$\begin{aligned}G(\theta,\lambda,t)&=G^{\text{Load}}(\theta,\lambda,t)+G^{\text{Rot}}(\theta,\lambda,t)\\&=\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}G_{lm}Y_l^m(\theta,\lambda)\end{aligned} \qquad [54a]$$

or

$$G(\theta,\lambda,t)=a\sum_{l=0}^{\infty}\sum_{m=0}^{+l}[C_{lm}\cos(m\lambda)+S_{lm}\sin(m\lambda)]P_{lm}(\cos\theta) \qquad [54b]$$

By investigating the first two terms (load and rot) of the geoid function, equivalent contributions to the Stokes coefficients $C_{21}^{\text{Load}}, C_{21}^{\text{Rot}}$, and $S_{21}^{\text{Load}}, S_{21}^{\text{Rot}}$ can be obtained. This simply involves multiplication of eqn [54a] by the spherical harmonic $Y_2^{1*}(\theta,\lambda)$, followed by the application of the orthogonality relation of the spherical harmonics themselves to determine a posteriori expressions for C_{21} and S_{21}. Since this procedure eliminates any influence of the space-independent mass conservation term from either of these expressions, insofar as these Stokes coefficients are concerned, this term has no influence.

We proceed, in what follows, to document the expressions for the two Stokes coefficients that are obtained by following this procedure and will deal separately with the contributions from the load and from the varying rotation, beginning with the former.

9.09.6.2.1 The load contribution

We obtain the load component of the space–time dependence of geoid height as

$$\begin{aligned}G^{\text{Load}}(\theta,\lambda,t)&=\int_{-\infty}^{t}d\tau\iint_{\Omega}\mathrm{d}\Omega'\left\{L(\theta',\lambda',\tau)G_G^L(\theta',t-\tau)\right\}\\&=\frac{a}{M_e}a^2\int_{-\infty}^{t}\mathrm{d}\tau\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\frac{4\pi}{(2l+1)}\sigma_{lm}(\tau)\\&\quad\left\{\delta(t-\tau)+k_l^L(t-\tau)\right\}Y_l^m(\Omega)\end{aligned} \qquad [55]$$

where we have employed the addition theorem for spherical harmonics (eqn [48]) and the orthogonality relation among spherical harmonics to eliminate the surface integral. Carrying the temporal convolution integral through this expression then delivers

$$\begin{aligned}G^{\text{Load}}(\theta,\lambda,t)&=\frac{1}{M_e}\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\frac{4\pi a^3}{(2l+1)}\\&\quad\{\nu_{lm}(t)+\int_{-\infty}^{t}\nu_{lm}(\tau)k_l^L(t-\tau)\}Y_l^m(\theta,\lambda)\end{aligned} \qquad [56]$$

We may extract the degree 2 and order 1 term from this equation by the application of the orthogonality condition once more, which leads to the following expression:

$$G_{21}^{\text{Load}}=\frac{1}{M_ea}\frac{4\pi a^4}{5}\left\{\nu_{21}+k_2^L*\nu_{21}\right\} \qquad [57]$$

Using the results in Section 9.09.3.5 transforms this expression into one involving the 'rigid' contributions to the moment of inertia tensor and thence to the total load components of the moment of inertia tensor as

$$\begin{aligned}G_{21}^{\text{Load}}&=\frac{-1}{M_ea}\sqrt{\frac{3}{10}}\left\{\left[I_{xz}^{\text{Rigid}}-iI_{yz}^{\text{Rigid}}\right]+k_2^L*\left[I_{xz}^{\text{Rigid}}-iI_{yz}^{\text{Rigid}}\right]\right\}\\&=\frac{-1}{M_ea}\sqrt{\frac{3}{10}}\left[I_{xz}^{\text{Load}}-iI_{yz}^{\text{Load}}\right]\end{aligned} \qquad [58]$$

Expressed in terms of the Stokes coefficients, this is equivalent to

$$\begin{aligned}\left[C_{21}^{\text{Load}},S_{21}^{\text{Load}}\right]&=g_1\left[\frac{I_{xz}^{\text{Load}}}{(C-A)},\frac{I_{yz}^{\text{Load}}}{(C-A)}\right]\ \text{where}\ g_1\\&=-\sqrt{\frac{3}{5}}\left[\frac{(C-A)}{M_ea^2}\right]\end{aligned} \qquad [59]$$

9.09.6.2.2 The rotational contribution

This is the most important part of this analysis. In parallel with eqn [55], we may write

$$G^{Rot}(\theta,\lambda,t) = \int_{-\infty}^{t} d\tau \iint_{\Omega} d\Omega'\ \Psi^{R}(\theta',\lambda',\tau)\ {}_{R}G_{G}^{T}(\phi,t-\tau) \quad [60]$$

where Ψ^R is defined in eqns [13]. We again invoke the addition theorem to expand the tidal Green function in reducing eqn [60], but now, the appearance of the renormalization factor in the definition of the Green function eliminates the factor $4\pi/(2l+1)$ that would otherwise appear. Taking this into account, eqn [60] may be expanded as

$$\begin{aligned} G_{21}^{\mathrm{Rot}} &= \frac{1}{g}\int_{-\infty}^{t}\left\{\delta(t-\tau)+k_2^T(t-\tau)\right\}\Psi_{21}(\tau)\mathrm{d}\tau \\ &= -\frac{\Omega^2a^2}{g}\sqrt{\frac{1}{30}}\left\{\left[m_x(t)-im_y(t)\right]+k_2^{T}*\left[m_x(t)-im_y(t)\right]\right\} \\ &= -\frac{\Omega^2a^2}{g}\sqrt{\frac{1}{30}}\left\{\frac{\left[I_{xz}^{\mathrm{Tot}}-iI_{yz}^{\mathrm{Tot}}\right]}{(C-A)}+k_f\frac{\left[I_{xz}^{\mathrm{Rot}}-iI_{yz}^{\mathrm{Rot}}\right]}{(C-A)}\right\} \end{aligned} \quad [61]$$

where Ψ_{21} is taken from eqn [13], and the results of Section 9.09.3.5 are used to obtain the final line in eqn [61]. Invoking the known relationships between the coefficients in the expansion of a field in terms of real spherical harmonics and those in the expansion of the same field in complex spherical harmonics, as in the preceding text, we have

$$\left[C_{21}^{\mathrm{Rot}},S_{21}^{\mathrm{Rot}}\right] = g_2\left[\frac{\left(I_{xz}^{\mathrm{Rot}}+I_{xz}^{\mathrm{Tot}}/k_f\right)}{(C-A)},\frac{\left(I_{yz}^{\mathrm{Rot}}+I_{yz}^{\mathrm{Tot}}/k_f\right)}{(C-A)}\right]$$

$$\text{where } g_2 = -\frac{\Omega^2 a k_f}{3g}\sqrt{\frac{3}{5}} \quad [62]$$

As discussed in Peltier et al. (2012), the $\delta(t-\tau)$ term in eqn [55] should not appear in eqn [61] if the purpose of the GIA-predicted geoid is to be compared with observations being provided by the GRACE satellites as an Earth-orbiting satellite makes measurements from an inertial frame of reference and is therefore incapable of observing an impact upon the geoid that is derivative of the direct effect of the inertial force that this term represents. However, since the physical geoid is related to sea level and since the ocean is attached to the rotating Earth, the water in the ocean basins does feel the centrifugal force and there is a component of the physical geoid that GRACE is unable to observe. As we will see, the fact that GRACE observes only an approximation to the geoid defined in terms of sea level introduces a difference between the inferred Stokes coefficients for the actual geoid such that the latter are larger than the former by a factor of ~2.

9.09.6.2.3 The total degree 2 and order 1 Stokes coefficients

Therefore, according to the SLE, $\dot{C}_{21},\dot{S}_{21}$ are related to the products of inertia in the following way:

$$\begin{aligned} \dot{C}_{21} &= g_1\left(\frac{I_{xz}^{\mathrm{Load}}}{(C-A)}\right)+g_1\left(\frac{I_{xz}^{\mathrm{Rot}}}{(C-A)}\right)+\frac{g_2}{k_f}\left(\frac{I_{xz}^{\mathrm{Tot}}}{(C-A)}\right) \\ \dot{S}_{21} &= g_1\left(\frac{I_{yz}^{\mathrm{Load}}}{(C-A)}\right)+g_2\left(\frac{I_{yz}^{\mathrm{Rot}}}{(C-A)}\right)+\frac{g_2}{k_f}\left(\frac{I_{yz}^{\mathrm{Tot}}}{(C-A)}\right) \end{aligned} \quad [63a,b]$$

where the constants of combination in the aforementioned expressions are

$$g_1 = -\sqrt{\frac{3}{5}}\frac{(C-A)}{M_e a^2}\quad g_2 = -\sqrt{\frac{1}{15}}\frac{a\Omega_o^2 k_f}{g} \quad [64]$$

The expression for the C_{21} coefficient that follows from the SLE-based geoid eqn [39] is then

$$C_{21} = \left(g_1+\frac{g_2}{k_f}\right)\left(\frac{I_{xz}^{\mathrm{Load}}}{(C-A)}\right)+\left(g_1+\frac{g_2}{k_f}\right)\left(\frac{I_{xz}^{\mathrm{Rot}}}{(C-A)}\right) \quad [65]$$

This expression may be reduced considerably by noting that the ratio g_1/g_2 is simply

$$\begin{aligned} \frac{g_1}{g_2} &= +\sqrt{\frac{3}{5}}\frac{(C-A)}{M_e a^2}\sqrt{\frac{15}{1}}\frac{g}{a\Omega^2 k_f} \text{ but } g=\frac{M_e G}{a^2} \\ \frac{g_1}{g_2} &= +3\frac{(C-A)}{M_e a^3\Omega^2}\frac{1}{k_f}\frac{M_e G}{a^2} \\ \text{and } k_f &= \frac{3G}{a^5\Omega^2}(C-A) \\ \text{so } \frac{g_1}{g_2} &= 1 \\ \text{and } C_{21} &= g_1\left(1+\frac{1}{k_f}\right)\left(\frac{I_{xz}^{\mathrm{Load}}}{(C-A)}+\frac{I_{xz}^{\mathrm{Rot}}}{(C-A)}\right) \end{aligned} \quad [66]$$

with a correspondingly simple result holding for S_{21}. Since the factor $(1+1/k_f)=2.06$ if k_f is taken to be ~0.94, this explains the flaw in the arguments of Chambers et al. (2010) as to the basis on which one should be able to determine the values of the degree 2 and order 1 Stokes coefficients directly from the time dependence of the products of inertia alone.

Insofar as model-predicted values of the individual Stokes coefficients based upon these formulae are concerned, these are as follows for the actual physical geoid when this is defined as the surface of constant gravitational potential that overlaps the surface of the sea in the absence of tides and currents. The results for the ICE-5G (VM2) model are

$$\left(\dot{C}_{21},\dot{S}_{21}\right) = (-0.50,+2.59)E-11\,\mathrm{year}^{-1} \quad [67]$$

Eliminating the $\delta(t-\tau)$ term in eqn [55] and therefore reducing the factor $1+1/k_f$ to unity, the values of these Stokes coefficients for the same model that GRACE would be expected to observe if the GIA process was the only process contributing to these coefficients become

$$\left(\dot{C}_{21},\dot{S}_{21}\right) = (-0.24,+1.24)E-11\,\mathrm{year}^{-1} \quad [68]$$

each of the coefficients therefore being further reduced by a factor of ~2.06. It is also worth noting the values for these Stokes coefficients quoted in the recent paper by Cheng et al. (2011) who acknowledged these as being provided by Wahr as

$$\left(\dot{C}_{21},\dot{S}_{21}\right) = (-0.2,+1.2)E-11\,\mathrm{year}^{-1} \quad [69]$$

These numbers are sufficiently close to those in eqn [37] to suggest that they have also been based upon use of the full ICE-5G (VM2) model of Peltier (2004). It is also important to note that the values being circulated are clearly not based upon the use of the formulation of the rotational response theory advocated in Mitrovica et al. (2005), which would reduce them to a further significant degree if it were assumed that the elastic thickness of a globe encircling lithosphere remained finite in

the limit of infinite time in response to tidal forcing. I view this modification to the 'traditional' form of the rotational response theory with skepticism for the reasons discussed in Peltier and Luthcke (2009) and Peltier et al. (2012).

It will be clear on the basis of the compilation of the GRACE-inferred values of the Stokes coefficients from **Table 2** of Peltier and Luthcke (2009) that neither of these sets of values is particularly close to the theoretically predicted values for the ICE-5G (VM2) model noted in the preceding text. Recapitulating the results listed in this table in Peltier and Luthcke (2009) for the purpose of comparison to those in the preceding text, we have

$$\begin{array}{lll} \text{GSFC}_\text{V}05 & \dot{C}_{21} = -1.65E-11\,\text{year}^{-1} & \dot{S}_{21} = +1.76E-11\,\text{year}^{-1} \\ \text{GSFC}_\text{V}06 & \dot{C}_{21} = -1.47E-11\,\text{year}^{-1} & \dot{S}_{21} = +1.76E-11\,\text{year}^{-1} \\ \text{GSFC}_\text{V}07 & \dot{C}_{21} = -1.68E-11\,\text{year}^{-1} & \dot{S}_{21} = +2.67E-11\,\text{year}^{-1} \\ \text{CSR} & \dot{C}_{21} = -1.80E-11\,\text{year}^{-1} & \dot{S}_{21} = +1.18E-11\,\text{year}^{-1} \\ \text{GFZ} & \dot{C}_{21} = -1.23E-11\,\text{year}^{-1} & \dot{S}_{21} = +0.28E-11\,\text{year}^{-1} \end{array} \quad [70]$$

where GSFC stands for the Goddard Space Flight Center and CSR stands for the Center for Space Research (both are US analysis centers) and GFZ stands for the GFZ German Research Centre in Potsdam. Comparison of these results based upon the 'observations' to those from either version of the theoretical model listed in eqns [36] and [37] will show that the differences are as large as those identified for the earlier less accurate version of the ICE-5G (VM2) model in Peltier and Luthcke (2009). That the accuracy of these GRACE inferences of the degree 2 and order 1 Stokes coefficients should be seen as highly compromised will be clear on the basis of the estimates that follow directly from the recent GRACE epoch analysis of the rotational observables published in Roy and Peltier (2011) that were discussed in **Section 9.09.4.3**. Given the results of this original analysis of the polar wander process in the post-1992 period, based upon the use of a modern Earth orientation data set, applying the results of that section, we obtain

$$\left(\dot{C}_{21}, \dot{S}_{21}\right) = (-0.40, +0.60)E-11\,\text{year}^{-1} \quad [71]$$

which we will take to be the most accurate observations available in what follows. These may be compared to the appropriate version of the model-based results in eqn [63] where it will be seen that there exist substantial differences. There is a factor of 2 difference in $\dot{C}_{21}$ with the observed value being higher than the model-predicted value. There is also a difference in the values of $\dot{S}_{21}$ with the model-predicted value being higher by a factor of 2 than the observed value. In Peltier and Luthcke (2009), it was hypothesized that discrepancies of this kind should be explicable in terms of modern land-ice melting, perhaps augmented (in our current view) by additional forcing due to changing storage of water on the continents. We turn to the construction of a preliminary test of the plausibility of this hypothesis in what follows.

9.09.6.3 GRACE Estimates of the Modern Rate of Land-Ice Melting

Support has recently been forthcoming for the plausibility of this hypothesis in the previously discussed paper by Roy and Peltier (2011). These analyses demonstrate that, beginning in the early 1990s, a decade prior to the launch of GRACE, both of the rotational observables began to deviate sharply from their previously established and Late Quaternary ice-age-controlled rates of secular variation. Of particular interest for present purposes is the fact that the observed rate of polar wander was significantly reduced and the direction shifted significantly to the east of the pre-1990s direction.

Although a refined forward model of the combined influence of the Late Quaternary ice-age cycle and modern land-ice melting will clearly be required to provide definitive proof or refutation of the land-ice melting hypothesis, we have constructed a preliminary forward model that includes the effect of modern land-ice melting and will employ this to fit the post-1992 rotational observations of Roy and Peltier (2011). In order to provide a clear demonstration of the quality of the model that has been most recently developed to serve as basis for such analysis, it is important to first document the fact that, in regions on the Earth's surface that are not influenced by modern land-ice melting but were strongly influenced by ancient ice age, the GIA model is able to make an accurate prediction of the modern hydrology-corrected GRACE observations. To this end, I show in **Figure 36(a)** comparison of the raw and surface hydrology-corrected Grace observations to the predictions of both the ICE-5G (VM2) model of Peltier (2004) and the newly developed ICE-6G (VM5a) model. Several important conclusions follow from the results shown in this figure. First, it will be noted that the correction of the raw GRACE data for the influence of surface hydrologic (groundwater storage) variations is important. Our implementation of this correction is based upon the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). This correction transforms the raw field into one characterized by a double-bull's-eye pattern with two extrema that straddle present-day Hudson Bay. Comparison of this corrected field to the predictions of the ICE-5G (VM2) model shows that, although the spatial location and amplitude of the GIA-predicted signal are reasonably accurate, this model does not predict the observed double-bull's-eye structure that is characteristic of the hydrology-corrected observations (a caveat is in order here in that the most recent analyses of the GRACE data discussed in Peltier et al. (2014) have shown that the absence of the double-bull's-eye structure from the raw GRACE data arises as a consequence of the application of the correlated error filter that was a necessary step in the data analysis procedure when only a few years of data were available. Given the extended number of years of data now available, this step is no longer necessary, and when it is eliminated, the two bull's-eyes appear in the raw data prior to the application of the hydrological correction). The predictions for the newly constructed ICE-6G (VM5a) model, however, successfully capture this double-bull's-eye structure, and so, this model is a marked improvement over ICE-5G (VM2). This model, the construction of which is fully described in Peltier et al. (2014), has been developed on the basis of the application of a refinement procedure based solely upon the use of GPS observations of vertical motion of the crust. In this procedure, the misfits of the GIA-predicted present-day rates of vertical motion to the GPS observations originally documented in Argus and Peltier (2010) were employed to adjust the thickness versus time history of the Laurentide ice sheet so as to eliminate the misfits to the observations that were characteristic of the ICE-5G (VM2) predictions. The GRACE observations were not themselves

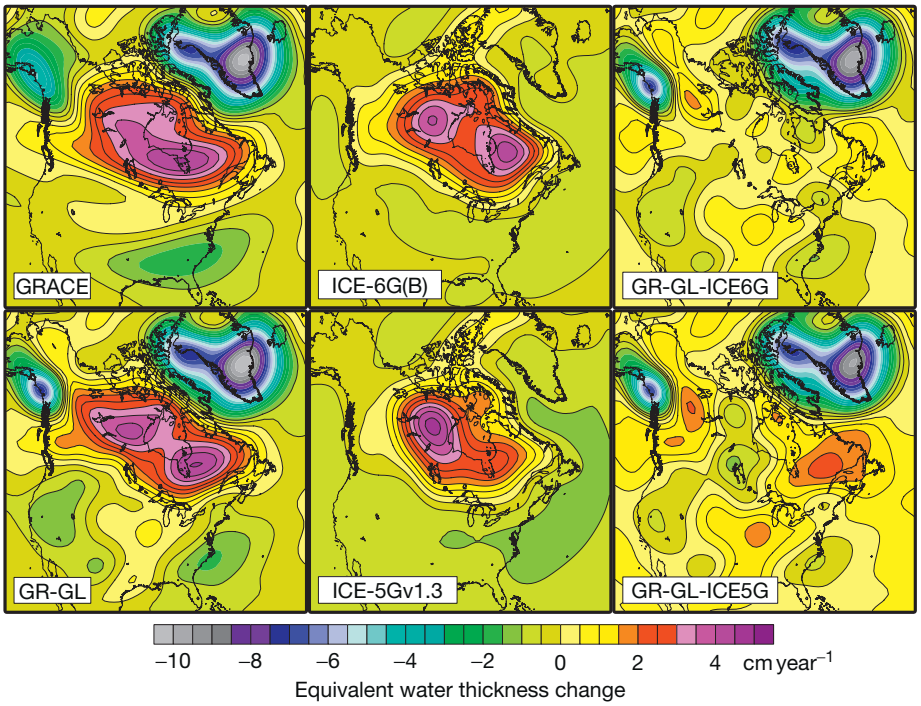


Figure 36 An analysis of the time-dependent gravity observations being provided by the GRACE satellites over the North American continent. The raw GRACE data are denoted 'GRACE,' while the result of filtering of these data to eliminate the influence of surface hydrology is denoted 'GRACE–GLDAS' in which GLDAS stands for the global surface hydrology data set of Rodell et al. (2004). GIA predictions are shown both for the ICE-5G (VM2) model and for the newly constructed ICE-6G (VM5a) model. The latter fields have been stripped of their degree 2 and order 1 Stokes coefficients as have the GRACE and GRACE–GLDAS fields so as to eliminate any ambiguity that might be attributed to the magnitude of the degree 2 and order 1 Stokes coefficients. Also shown are the differences between the two GIA model predictions and the GRACE–GLDAS field. The latter figures illustrate the extent to which the GIA models are able to annihilate the hydrology-corrected GRACE observations. Evident is the fact that the ICE-6G (VM5a) model provides a much improved fit to the observations as it fully explains the 'double-bull's-eye' form of the observations, whereas the ICE-5G (VM2) model does not. Over North America, the adjustment of the Laurentide ice loading history was adjusted to eliminate the misfits to the GPS observations of the ICE-5G (VM2) model identified in Argus and Peltier (2010). The differences between the hydrology-corrected GRACE field and the GIA models isolate the signals over both Greenland and Alaska that are related to global warming-induced land-ice melting.

employed to refine the model. Since the Laurentide ice-sheet complex was by far the most significant at the LGM, the fact that the new ICE-6G (VM5a) model provides such an excellent fit to these observations must be seen to provide a useful independent check on its quality. It should also be noted that in this comparison, the degree 2 and order 0 and the degree 2 and order 1 geoid coefficients are included but the observed coefficients are taken to be equal to those inferred on the basis of the Roy and Peltier (2011) analyses.

Table 2, the first two rows of which were previously discussed in Section 9.09.4.3, lists results for the rotational observables for a series of models of glaciation history including model ICE-6G (VM5aD1) of Peltier and Drummond (2010), which contains no modern land-ice melting and which is one of the models employed for the purpose of previous discussion in this chapter. This model reasonably well fits the pre-1992 values of the nontidal acceleration and the speed and direction of polar wander as previously commented. However, as also noted in the table, it significantly misfits the degree 2 and order 1 Stokes coefficient $\dot{S}_{21}$.although the result for the $\dot{C}_{21}$ coefficient is close to that observed according to the Roy–Peltier inference that is listed in the last row of the table based upon the discussion in the previous subsection. The next model for which results are also listed is for a model denoted 'A' in which only the influence of land-ice melting at the poles is included. These polar sites of regional melting include Alaska, Antarctica, and Greenland, with the rates of melting in each of these regions fixed to those inferred by Peltier (2009) on the basis of GRACE observations. These melting rates (in global sea-level rise equivalents) are 0.12 mm year^{-1} for Alaska, 0.34 mm year^{-1} for Antarctica, and 0.59 mm year^{-1} for Greenland. The influence of modern land-ice melting is added to model ICE-6G (VM5aD1), which includes only the influence of the Late Quaternary ice-age cycle. Inspection of the results for the complete set of rotational observables in this table demonstrates that neither the polar wander speed and direction, nor the $\dot{J}_2$ observation, nor the Stokes coefficients are well fit by this model. This is in spite of the fact that, if one were to

focus solely upon the impact on the nontidal acceleration-related $\dot{J}_2$ parameter and consider only the perturbation to the inferred pre-1992 rate of Roy and Peltier (2011), one would obtain the result shown in the previously discussed **Figure 28** for the perturbation due to the collective effect of all three regions of land-ice melting. Results similar to those in **Figure 28** were recently published in Nerem and Wahr (2011), who apparently believed these to fully explain the shift in the $\dot{J}_2$ observable first identified and ascribed to land-ice melting in Roy and Peltier (2011); but see also Cox and Chao (2002) and Dickey et al (2002). What Nerem and Wahr seem to have missed, however, is that when the influence is added to that of a model that reasonably well fits the pre-1992 values for the rotational observables, the misfits to polar wander speed and direction thereby induced are unacceptable. This fact was fully discussed previously but it is worth reiterating here. Especially important for present purposes, however, is the further increase in the misfits to the Stokes coefficients associated with the parent ice-age-only model ICE-6G (VM5aD1). The observed post-1992 shift in the secular rate of change in the nontidal acceleration cannot therefore be explained by the melting of polar ice from Greenland, Alaska, and Antarctica alone, as the other rotational constraints are then violated.

It is important to note that in these calculations, the full gravitationally self-consistent form of the SLE including rotational feedback has been employed to compute the impact upon all of the rotational observables. Figure 3 of Peltier et al. (2012) illustrates the most important consequence of polar land-ice melting in this regard, in terms of what might be referred to as the 'opposite hemisphere effect.' This figure from that paper shows the results for the sea-level rise that would be caused by the melting of all of the ice on either Greenland or West Antarctica over a period of a century. Notable is the fact that sea-level falls in the immediate vicinity of the land mass from which ice has disappeared and rises dramatically in the opposite hemisphere. This effect was first explored in some of the earliest work involving the use of the SLE formalism (Clark, 1976; Clark and Lingle, 1977). It is crucial to understanding the sea-level teleconnection involved in which the meltwater pulse 1B event due to melting of land ice from Antarctica causes a nonmonotonic rise of sea level at sites once covered by the Scottish ice-sheet complex in the opposite hemisphere.

Returning now to the discussion of the data in **Table 2**, this table also shows results for eight additional models, the first four of which are denoted B, C, D, and E. For these models, additional sources of land-ice melting have been included in the calculation for three of the primary regions in which this process is occurring according to the tabulation in Dyurgerov (2010). These three locations include the mountains of South America, the Himalayan mountains of Asia, and the mountains of western North America, which appear clearly defined in **Figure 37** where they are seen to be the regions of highest topographic relief on the planet. The high topography melt rates assumed for these models are based upon simple scans of hypothetical rates of melting at each of these additional locations in such a way as to test whether it would be possible to fit the new values for the rotational observables inferred by Roy and Peltier (2011) by appropriate choices for them. As shown in the table for models B, C, D, and E, these values of the post-1992 rotational observables are indeed somewhat better fit by these models. Furthermore, as inspection of the results in the table will show, the misfits to the Stokes coefficients estimated on the basis of Roy–Peltier data have also been reduced. The results for the final set of four models in the table are for a simple variation upon models B, C, D, and E in which the radial viscosity profile VM5aD1 is simply replaced by viscosity profile VM5a, which is a simple multilayer approximation to model VM2. For these additional models, the rates of melting are kept fixed to the values assumed for models B, C, D, and E, which are, for the mountains of South America, the Himalayas, and western North America, in global sea-level rise equivalent terms, respectively, 0.72, 0.42, and 0.45 mm year^{-1} for model B; 0.72, 0.37, and 0.47 mm year^{-1} for model C; 0.67, 0.47, and

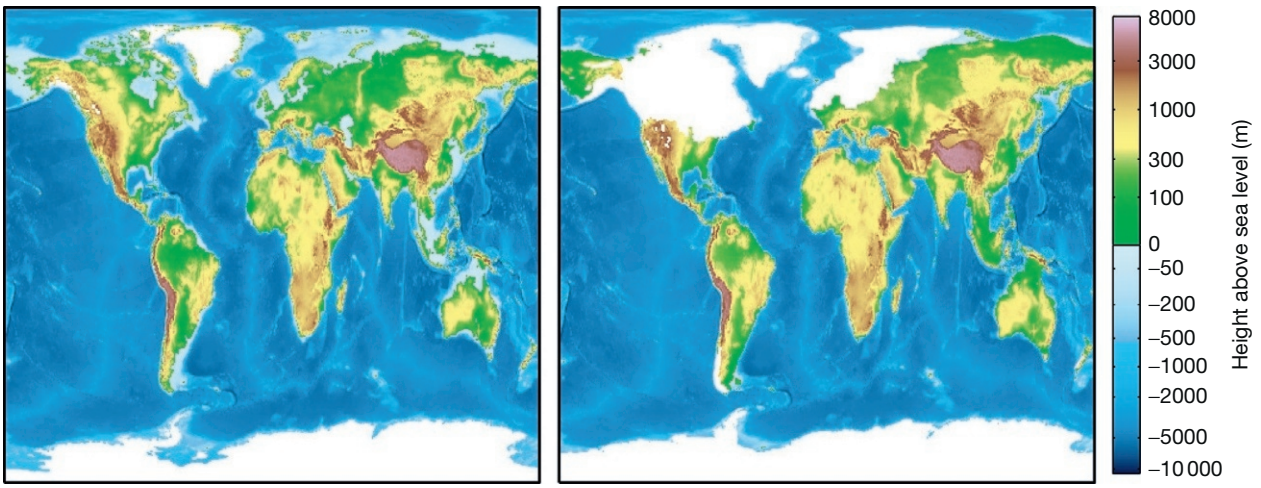


Figure 37 Comparison of surface topography of the non-ice-covered regions of the Earth's surface under modern and LGM surface conditions. Once more, the oceans are depicted in terms of the depth-averaged Brunt–Väisälä frequency. The highest topography regions in this figure, namely, the Himalayas, the Andes, and the North American Cordillera, are the locations at which it is assumed that significant land-ice melting is occurring for the purpose of analyzing the plausibility of the modern land-ice melting hypothesis of Peltier and Luthcke (2009) as an explanation of the rotational data of Roy and Peltier (2011).

0.49 mm year^{-1} for model D; and finally 0.65, 0.47, and 0.50 mm year^{-1} for model E. The final set of four models clearly deliver a very significant further improvement of the quality of fit to the rotational observables and therefore further reduce the misfits to the Roy and Peltier inferred values of the Stokes coefficients. For several of these additional models, the fits are in fact excellent. It is extremely important to note that the total rates of land-ice melting, in eustatic sea-level rise equivalent terms, for models B, C, D, and E, when the contributions from Greenland, Antarctica, and Alaska are included, are respectively, 2.63, 2.60, 2.63, and 2.67 mm year^{-1}. These net rates are in reasonable accord with the constraint on the net rate of global sea-level rise measured by the altimetric satellites Topex/Poseidon and Jason-1/-2 when further adjusted for the steric contribution, a contribution that Cazenave et al. (2009) had inferred to be ~0.37 mm year^{-1} for the GRACE era based upon their analysis of ARGO float data. Church et al (2011), however, had recently quoted a much larger preferred value for the steric contribution of 0.88 ± 0.33 mm year^{-1} and a significantly higher value for the rate of absolute sea-level rise when tide gauge data are employed to augment the basic altimetric observations. It is clear that no consensus has yet emerged on the important issue of sea-level rise budget closure.

My intention in presenting the results of the previously described simple models, based upon the original discussion in Peltier et al. (2012), is nevertheless not to present a fully satisfactory refined model for the melting of small glaciers and ice caps (GICs) but simply to demonstrate that the physical impact of this process could not be ruled out as a possible explanation of any misfit of the ice-age-only model of continental deglaciation to the geoid Stokes coefficients of degree 2 and order 1. Of course, this specific model might be thought to be entirely ruled out based upon the very large geoid anomalies that it predicts over the mountainous regions (not shown), which have no counterpart in the GRACE-observed field of geoid height time dependence. Whether these geoid anomalies could be masked to some (perhaps significant) degree by the continuing uplift of the mountains is an issue that remains to be adequately explored. Furthermore, there clearly exist additional regions of land-ice melting that must be contributing to the current input of freshwater into the global oceans, and their contributions have yet to be inferred with sufficient accuracy. In the next section of this chapter, we shift our focus to additional aspects of the Earth's rotational history, aspects that involve not the planets intrinsic rotation but rather its orbital rotation around the Sun.

9.09.7 The Impact of Variations in the Geometry of the Earth's Orbit Around the Sun upon Climate System Evolution

In the previous sections of this chapter, the focus has been fixed upon the history of the Earth's rotation as observed in the body-fixed coordinate system, the terrestrial fame, in which the anomalies in the variations in the l.o.d. and polar motion are most usefully described. Equally important aspects of the evolution of Earth's rotation, however, are those that are most usefully described in the celestial frame provided by the field of stars of our own and more distant galaxies. In this celestial frame of reference, the orbit of the Earth in its rotation around the Sun is observed to be characterized by subtle time dependence in its geometry. These variations are due to the action of gravitational n-body effects in the solar system that cause the ellipticity of the otherwise Keplerian elliptical orbit to vary with time and similarly impose subtle time dependence upon orbital obliquity, that is to say upon the angle that the spin axis makes with the plane of the ecliptic (see Figure 38). As it happens, this aspect of the history of the Earth's rotation exerts a profound effect upon low-frequency climate variability and in particular is the ultimate cause of the Late Quaternary ice-age cycle that has been the focus in several of the preceding sections of this chapter. This so-called astronomical theory of the ice ages, due originally to Milankovitch (e.g., 1940), was finally verified as essentially correct in the seminal paper by Hays et al. (1976) that was based upon the analysis of oxygen-isotopic data from deep-sea sedimentary cores such as the ODP Site 677 data from the Panama Basin shown previously in Figure 5 and which was analyzed in detail in Shackleton et al. (1990). The origins of the particular time series shown in this figure are in fact extremely important, from the perspective of both our understanding of the Late Pleistocene ice-age cycle and our appreciation of the crucial role that the history of the Earth's rotation has played in our understanding of the detailed characteristics of the process of plate tectonics itself.

Prior to the analysis presented in the Shackleton et al. paper, it had been widely believed, based upon Potassium–Argon dating of the age of the most recent Brunhes–Matuyama reversal of the polarity of Earth's magnetic field, that this event had occurred ~730 000 years before present. This assumed age was, until the appearance of this paper, universally employed in the construction of the chronology of all of the deep-sea

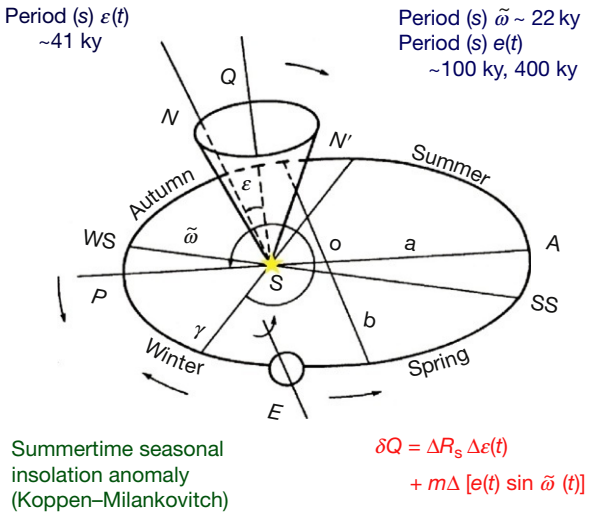


Figure 38 Geometry of the Earth's orbit around the Sun depicting the present positions of the apogee and the perigee, as well as the summer and winter solstices and the spring and autumn equinoxes. The tilt of the spin axis with respect to the plane of the ecliptic is represented by the parameter 'ε,' whereas the eccentricity of the orbit is represented by the parameter 'e.' As a consequence of the influence of the precession of the spin axis in space, the position of the equinoxes and solstices rotates around the orbital ellipse with a period of ~22 000 years.

sedimentary cores that were employed as constraints upon the understanding of climate system processes over the Pleistocene epoch. This date was also employed to constrain the speeds with which oceanic plates were assumed to 'spread' away from oceanic ridge crests in response to the underlying mantle convection process. The history of the process that led to our present understanding that the age of this critical control point for planetary chronology was 780 000 years rather than 730 000 years is important from the perspective of our appreciation of the role that the Earth's rotation history plays in our understanding of the system and so is worth reviewing in some detail.

9.09.7.1 The Astronomical Imprint on Oxygen-Isotopic Records from Deep-Sea Sedimentary Cores

Basic to this understanding is an appreciation of the extent to which modern astrophysical theory has been successful in the reconstruction of the evolving geometry of the Earth's orbit around the Sun over the past several million years of Earth history. Early work on this problem was reviewed and significantly extended by Laskar (1988) who employed a mixture of analytic and numerical techniques to demonstrate that it was possible to perform stable integrations of the geometry of the orbit that extended in excess of several million years into the past. These reconstructions were tested by performing brute force initial value integrations, starting from a modern ephemeris, of the system of gravitational n-body equations governing the evolution of the entire solar system, a system of equations that included an accounting for tidal interactions among the planets and their satellites and for relativistic effects. The results of such brute force integrations were first described in an important paper by Quinn et al. (1991). These analyses demonstrated that the solution of Laskar (1988) was highly accurate and could therefore be employed with confidence to accurately predict the variation of the strength of the solar insolation received by the Earth in the course of its orbit around the Sun as a function of time and as a function of latitude. Earlier predictions of these variations produced by Berger (e.g., 1978) were thereby shown to become inaccurate prior to ~ 1 million years before present. The numerical reconstructions of the evolving geometry of the Earth's orbit consist of predicted time series for the evolution of orbital ellipticity and orbital obliquity. From these quantities, one may compute the latitude-dependent summertime seasonal anomaly in solar insolation received at the top of the atmosphere that is presumed to drive the glaciation–deglaciation process according to the Koppen–Milankovitch hypothesis on the basis of the expression

$$\delta Q(\theta, t) = \Delta R_s(\theta)\Delta\varepsilon(t) + m(\theta)\Delta\left[e(t)\sin\left(\tilde{\omega}t\right)\right] \qquad [72]$$

in which $\delta Q(t)$ is the deviation, in watts per square meter, of the insolation received at the surface of the Earth during the summer season relative to the average of summer insolation over, say, the timescale of the entire 900 000-year period during which the 100 000-year ice-age cycle has been an active element of climate system evolution. The first recognition of the existence of the essentially 100 000 year periodic ice-age cycle was recorded in the paper of Broecker and van Donk (1970). The latitude dependence of the summertime seasonal insolation anomaly is introduced through the quantities $\Delta R_s(\theta)$ and $m(\theta)$, whereas the time dependence is governed by the variation in orbital obliquity $\Delta\varepsilon(t)$ and time-dependent eccentricity of the orbit $e(t)$, which enters the expression for the insolation anomaly only through a modulation of the influence of orbital precession, which occurs at an angular frequency $\tilde{\omega}$ due to the same oblateness of the figure of the Earth that is modulated so effectively by the ice-age cycle to produce variations in the l.o.d. This phenomenon of orbital precession is caused by the action of the gravitational field of the Sun upon the oblate form of the planet whose spin axis is tilted with respect to the ecliptic plane. Figure 39 illustrates the variation with time of the individual elements that appear in eqn [72] in the preceding text, whereas Figure 40 shows the actual distribution of the received insolation over the surface of the Earth as a function of latitude and time of year, both for the present day and for the deviations from this distribution at several fiducial times over the most recent 100 000-year cycle of Late Pleistocene glaciation. Inspection of the distribution of the anomaly for 116 000 years before present at the onset of the most recent of these cycles of continental ice-sheet expansion demonstrates that this event was in fact apparently triggered by a high-latitude summertime seasonal insolation anomaly of magnitude in excess of 60 $\mathrm{W\,m^{-2}}$, in accord with expectations based upon the Milankovitch hypothesis. At LGM itself, the deviation of the insolation regime from that characteristic of present climate was actually rather small. During the mid-Holocene interval centered upon 6 ka, the nature of the insolation forcing was such that the northern-hemisphere continents would have been anomalously warm during the summer season, thus explaining the increased intensity of the summer monsoon circulations that were characteristic of the atmospheric general circulation at that time. Figure 41 compares the power spectrum of the oxygen isotope anomaly from deep-sea core ODP Site 677 to the power spectrum of the summertime seasonal insolation anomaly at 60° N latitude for both the first and last million years of the Pleistocene epoch. Inspection of this intercomparison will show that the power in the insolation forcing at the 100 000-year period on which the continental ice sheets expand and contract is negligible. The implications of this comparison are profound as it provides a direct demonstration of the intensity of the nonlinearity of the climate system that is involved in converting the incoming insolation signature into the climate response. It is only through the action of such nonlinearity that the system may feel the effect of the 100 000-year modulation of the influence upon received insolation due to the effect of orbital precession that arises from the variation of orbital eccentricity. Precisely what the origin is of this nonlinearity remains a subject of intense debate, but it may be connected to the cause of the covariation of atmospheric carbon dioxide with continental ice volume that is so clearly revealed in the Vostok ice core from the continent of Antarctica (e.g., see Jansen et al. (2007) for a detailed discussion of the ice core records and Shackleton (2000) for a discussion of the phase relationships between the atmospheric carbon dioxide record and the ice-volume record.

9.09.7.2 'Orbital Tuning' and the Age of the Brunhes–Matuyama Geomagnetic Polarity Transition

Given this background, we may proceed to consider the analyses that led to the redating of the oxygen-isotopic record from ODP Site 677 that has played such an

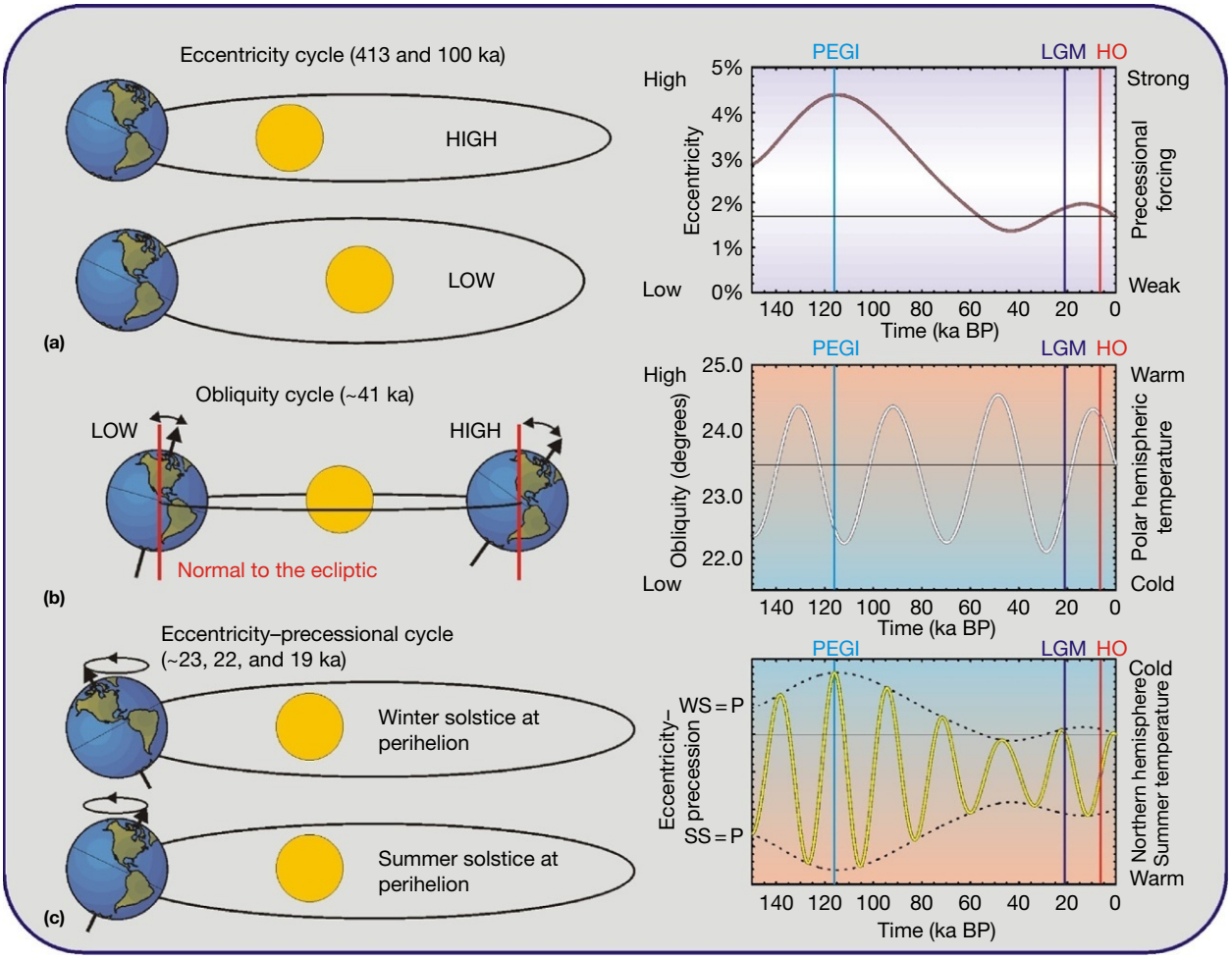


Figure 39 Evolution with time, over the most recent 100 000-year cycle of Late Pleistocene climate variability, of the parameters that control the variation of the solar insolation received at the top of the atmosphere.

important role in revising our understanding of the Pleistocene timescale. Figure 43 shows the original depth versus timescale that was originally ascribed to this core by Shackleton (personal communication in 1988 when the data were passed to the present author for analysis). Notable is the 'dogleg' in this relationship that is apparently near the then assumed age of 730 000 years of the Brunhes–Matuyama transition that was widely accepted as correct at that time. Based upon the assumption of the validity of the Milankovitch hypothesis, the depth–age relationship for this core was refined in Toronto by applying the following strategy. First, the relationship between depth in the core and calendar year age was refined by applying a band pass filter to the isotopic time series with a central frequency equal to that of the obliquity timescale of 41 000 years, and the depth–time relationship was adjusted iteratively so as to maximize the coherence between the astronomical forcing at this period and the ice-volume response. Next, the same process was applied to the variability in the eccentricity–precession band in which the modulation of the precession effect due to the eccentricity variation occurs. Once more, the age–depth relationship was adjusted in order to optimize the coherence between astronomical signal and ice-volume response. The frequency-domain form of the band pass filters employed in this analysis together with the power spectrum of the oxygen isotope time series is shown in Figure 42. The process of refinement of the age–depth relationship was continued by returning to the obliquity band and then returning to the eccentricity–precession band until an entirely stable result was obtained. The new age–depth relationship for the 677 core is also shown in Figure 42 as the dashed line labeled model II where it is observed to correspond to a very nearly linear relationship between depth and age, suggesting that in fact, the sedimentation rate at this site had been very nearly constant throughout the entire Pleistocene interval. The initial depth–age model provided to the author by N. J. Shackleton is shown as model IV. The models denoted I and III on the figure are intermediate results obtained in the iterative sequence of steps in the 'orbital tuning' process. The quality of the fits between the astronomical forcing and ice-volume response records in the two pass bands employed in the application of this methodology is shown in Figure 44(a) and 44(b), respectively, where a very nearly precise agreement was obtained (see Peltier (1996a) in which the

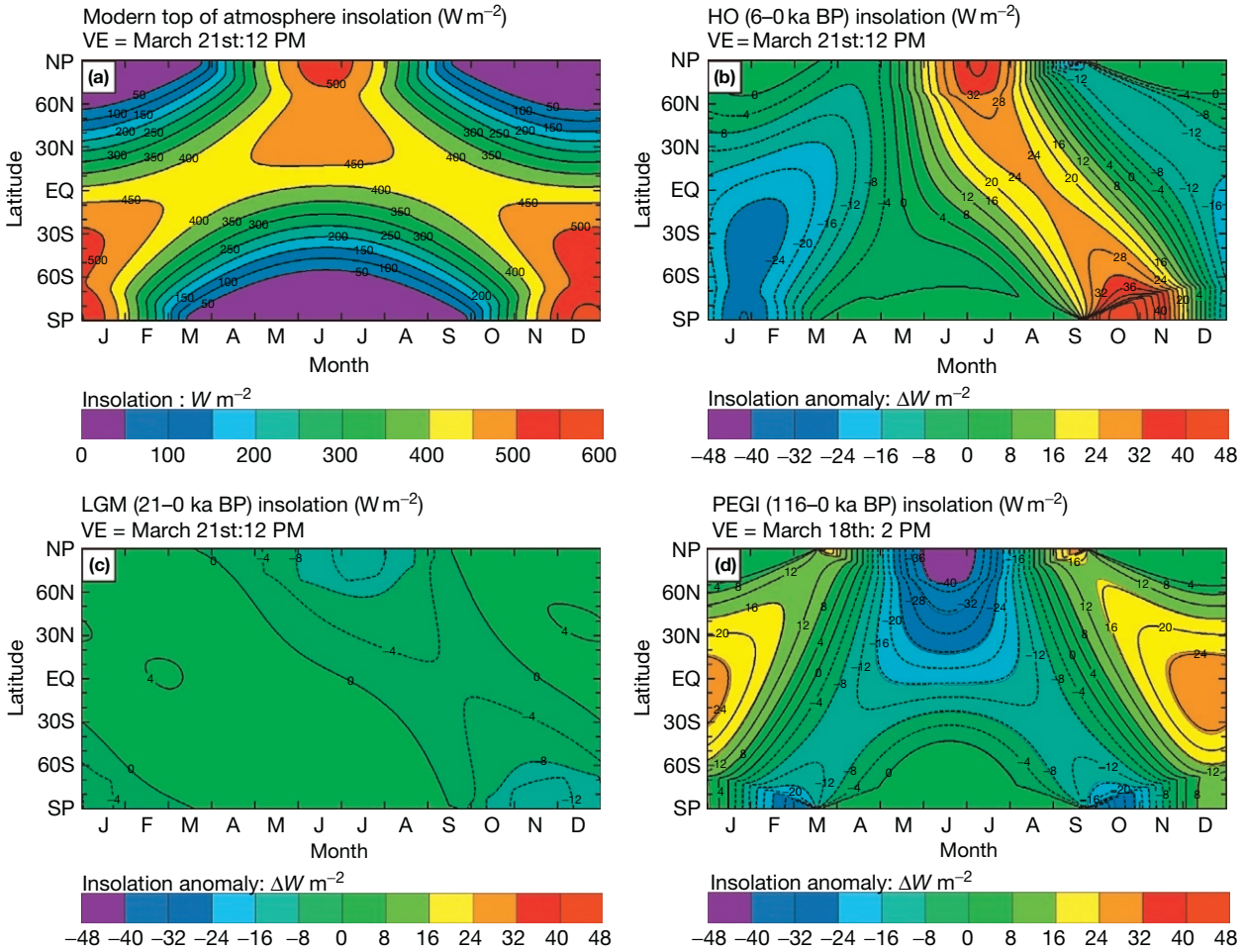

Modern top of atmosphere insolation (W m−2)
VE = March 21st:12 PM
HO (6–0 ka BP) insolation (W m−2)
VE = March 21st:12 PM
LGM (21–0 ka BP) insolation (W m−2)
VE = March 21st:12 PM
PEGI (116–0 ka BP) insolation (W m−2)
VE = March 18th: 2 PM
Latitude
NP
60N
30N
EQ
30S
60S
SP
J F M A M J J A S O N D
Month
Insolation : W m−2
0 100 200 300 400 500 600
Insolation anomaly: ΔW m−2
−48 −40 −32 −24 −16 −8 0 8 16 24 32 40 48

Figure 40 Distribution of solar insolation received at the top of the atmosphere as a function of time of year and latitude for three different epochs. (a) shows the present-day distribution, whereas (b)–(d) respectively show the deviation from the modern distribution for the mid-Holocene warm period at 6000 years before present, for Last Glacial Maximum at ~21 000 years before present, and for the end of the Eemian interglacial at 116 000 years before present, the time during which the extremely strong negative summertime seasonal insolation anomaly existed that was responsible for the onset of the most recent 100 000-year glacial cycle.

original results that led to the paper by Shackleton et al. (1990) were eventually first published).

This analysis suggested that there must have been an error in the previously accepted age of the Brunhes–Matuyama transition in the polarity of the Earth's magnetic field, with the actual age of the transition being very close to 780 000 years rather than to the originally assumed age of 730 000 years, a difference of ~7% or 50 000 years. In the paper by Shackleton et al. (1990), the initial Toronto result was checked against a large number of other oxygen-isotopic records from deep-sea cores and found to be reproducible in every instance. Very soon after the appearance of the Shackleton et al. publication and in response to it, the age of the Brunhes–Matuyama transition was redated using the argon 39–argon 40 method by Baksi (1992) whose analysis immediately confirmed the validity of the new age for this most recent polarity transition in the Earth's magnetic field, an age that had been determined by the application of the 'orbital tuning' methodology. It should be clear that this combination of results has served to fully verify the validity of our present understanding of the Late Pleistocene ice-age cycle as being caused by the history of the variations in the geometry of the Earth's rotation around the Sun.

9.09.8 The Earth's Rotation Variations, Mantle Convective Mixing, and Plate Tectonics-Related Processes

On the ultra-long timescales on which the planet evolves thermally, it remains very useful to continue to discuss the variations in the Earth's intrinsic rotational state in terms of variations in the l.o.d and variations of the position of the poles with respect to a body-fixed coordinate system. On such long timescales, the mantle convection process leads to a continuous reorganization of the land-sea distribution as the continents are caused to 'drift' over the surface of the Earth. Because this drift is associated with changes in all of the elements of the moment of inertia tensor of the planet, in consequence, a reaction of the rotational state in both l.o.d. and polar motion

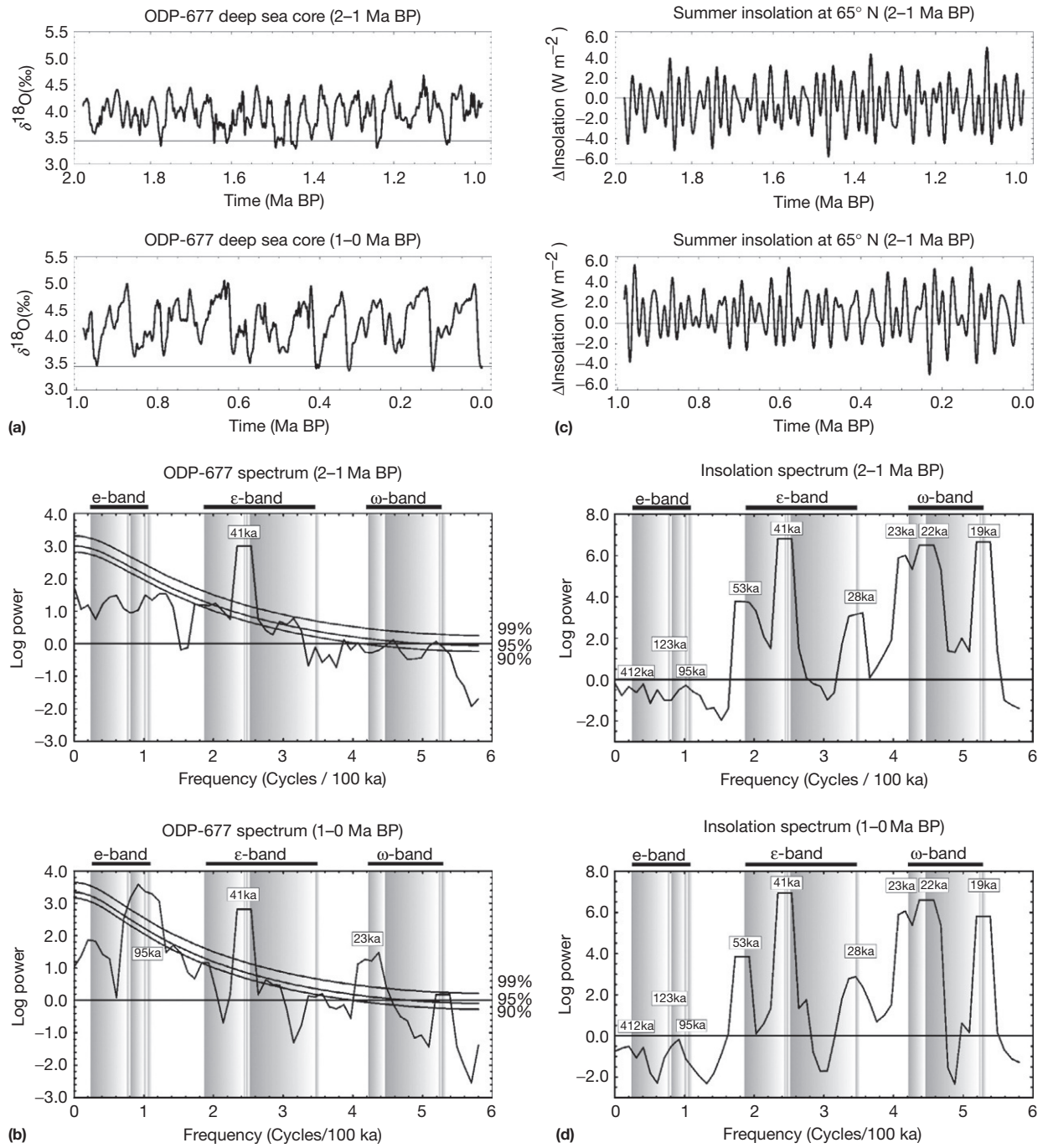


Figure 41 Intercomparison of the oxygen-isotopic time series from ODP Site 677 and its power spectrum to the time series and power spectrum of the solar insolation forcing for both the first and last million years of these records. Notable is the extremely weak insolation forcing at the period of 100 000 years, which is the period at which the climate system response is strongest.

is produced. At least equally important from the former perspective, however, are the changes in the l.o.d. that occur as an indirect consequence of the continuous rearrangement of the continent ocean distribution. This additional physical effect acts through the dissipation of the tides raised in the oceans, primarily by the gravitational attraction of the Moon, which is the origin of the tidal deceleration of rotation previously discussed. In what follows, I will provide a brief discussion of the current state of understanding of each of these additional sources of rotational excitation, beginning with the latter.

9.09.8.1 Tidal Despinning of the Earth's Rate of Axial Rotation: Tidal Rhythmites and the l.o.d.

Although space limitations mitigate against providing a detailed review of this important very long-timescale

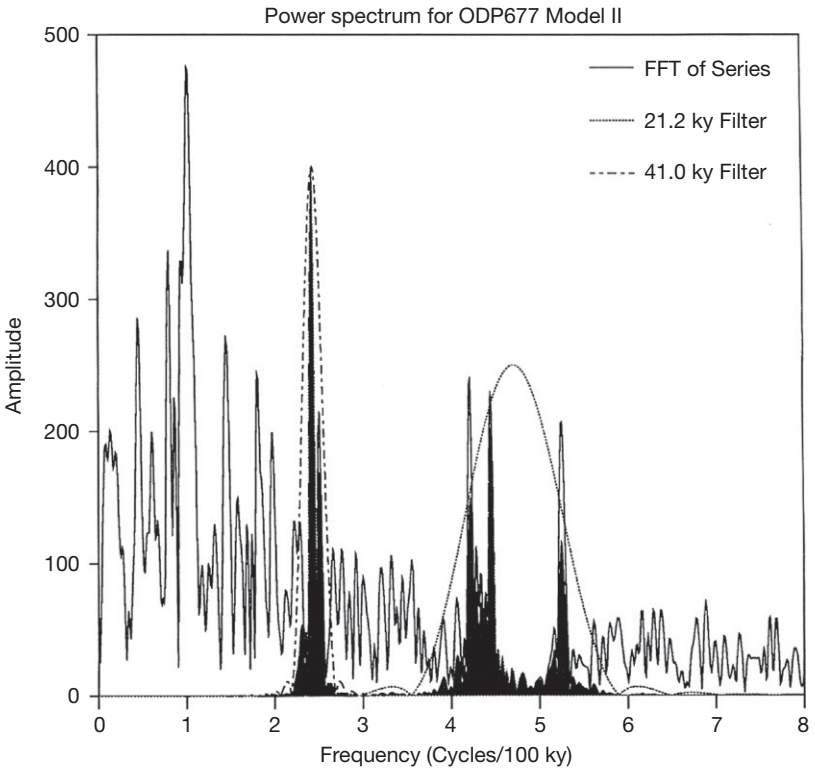


Figure 42 Frequency-domain forms of the band pass filters employed in the Toronto application of the orbital tuning methodology to the $\delta^{18}O$ record from ODP Site 677.

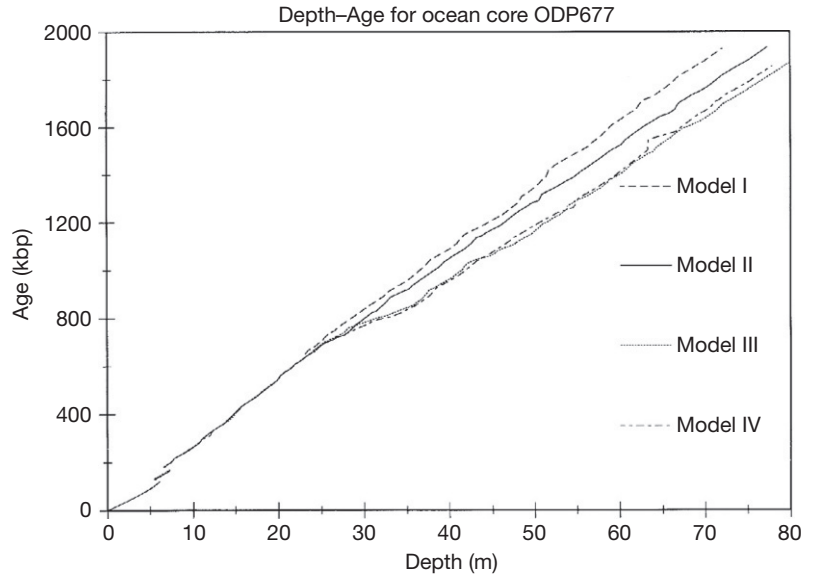


Figure 43 Oxygen-isotopic record from ODP Site 677 as originally provided by N. J. Shackleton to the writer in 1988 on the depth versus timescale constrained to fit the assumed age of 730 000 years of the Brunhes–Matuyama geomagnetic reversal. Notable in the original record is the 'dogleg' that was forced to occur in the depth–time relationship by virtue of this assumption. Also shown in the figure are the revised depth–time relationships deduced by 'orbital tuning' of this record that was accomplished at the University of Toronto and that led to the prediction of the necessity of revising the age of this geomagnetic reversal.

mechanism through which highly significant secular changes in the l.o.d. occur, namely, through the action of ocean tidal friction alone, it will be important to at least record the fact that such changes are large and that there exist extremely useful observations that can be brought to bear upon the problem. Although the problem of the ocean tides has a long and distinguished history of investigation in the geophysical sciences, it remains an active area of investigation as a consequence of

the need to develop a much more detailed understanding of tidal dissipation. This dissipation acts through two distinct mechanisms, namely, through bottom friction and through the dissipation associated with the breaking of the internal tide that is generated by the interaction of the barotropic tide with bottom topography in the stably stratified abyssal ocean (e.g., see Griffiths and Peltier (2008, 2009) and Salehipour et al (2013)). As a consequence of the action of both contributions to the net dissipation, the tidal bulge in the oceans raised by the gravitational attraction of the Moon is phase-delayed, and so, the spin angular momentum of the planet is subjected to a net torque, which decreases the rotation rate (see **Figure 45**).

Because of the dominant control that the semidiurnal M_2 tide in the oceans exerts upon coastal marine sedimentologic processes, it has been possible on the basis of the study of 'tidal rhythmites' to infer what the l.o.d. has been at certain epochs in the distant past. For example, in Williams (1997, 1998, 2000), analyses are described of South Australian sequences for the Late Neoproterozoic (~620 million years before present, near the time of the Marinoan glaciation) that indicate a rotational state characterized by 13.1 ± 0.1 synodic (lunar) months per year, 400 ± 7 solar days per year, and a l.o.d. of 21.9 ± 0.4 h. **Figure 46**, from Williams (1998), shows an example of the subtle banding characteristic of the rhythmites of the Elatina Formation. He also inferred on the basis of these data a mean rate of lunar recession since that time of 2.17 ± 0.31 cm year^{-1}, slightly more than half the present-day rate of lunar recession of 3.82 ± 0.07 cm year^{-1} obtained on the basis of lunar laser ranging, as discussed previously in **Section 9.09.4.1** of this chapter. **Figure 47** illustrates the manner in which the sedimentary sequences in the rhythmite may be translated into information that can be interpreted quantitatively to infer these numerical results. Similar data from earlier epochs are also discussed in Williams (2000) review paper and are illustrated in **Figure 48**, which suggest that a close approach of the Moon to the Earth would not have occurred during the entire Proterozoic aeon (2450–620 Ma). These results are clearly important for our understanding of the origin of the Moon and of the processes that control tidal friction. The current configuration of the ocean basins is such

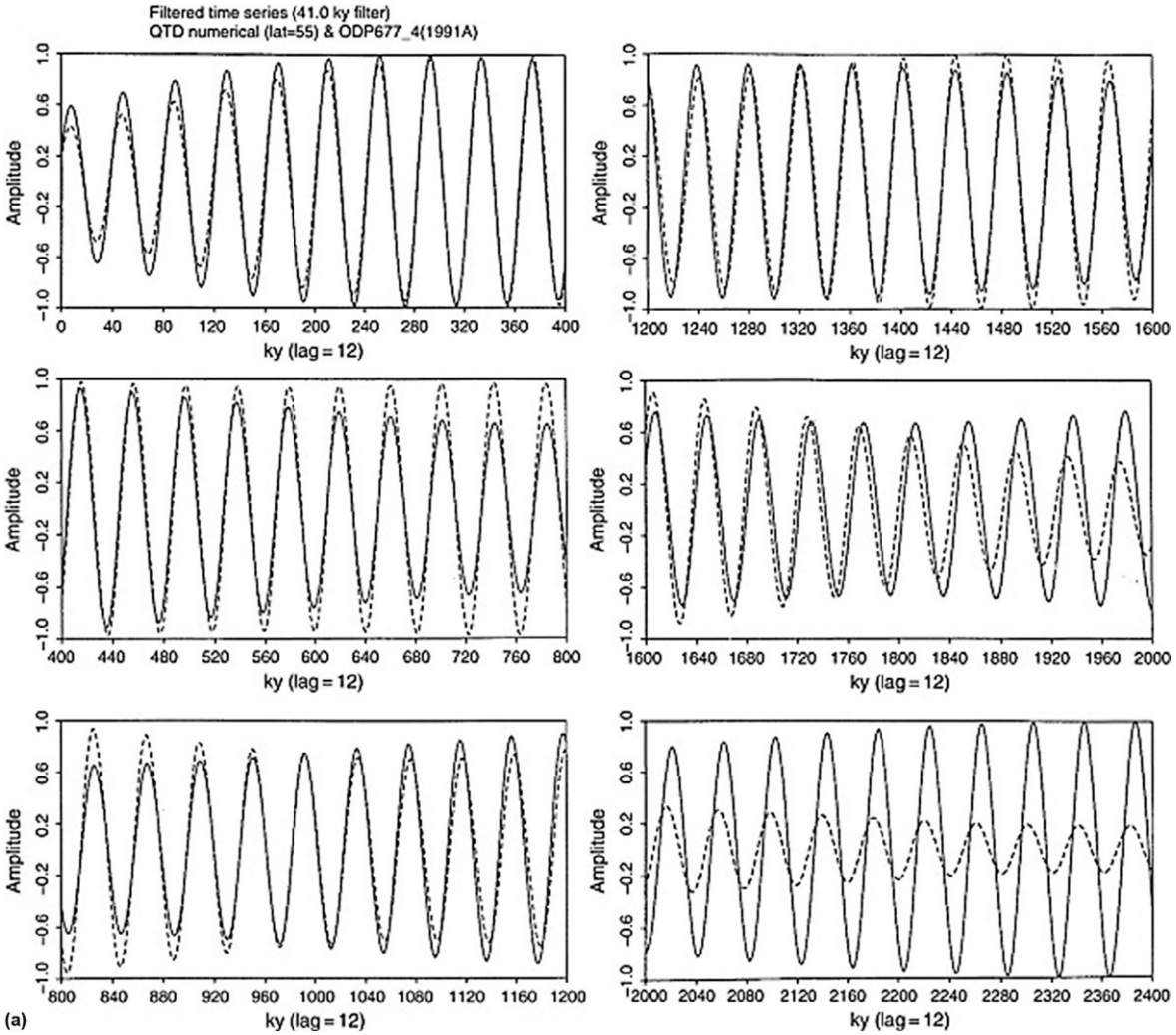


Figure 44 (a) Coherence between the astronomical forcing at the period of 41 000 years and the oxygen-isotopic variability at the same period after orbital tuning at the University of Toronto.

(Continued)

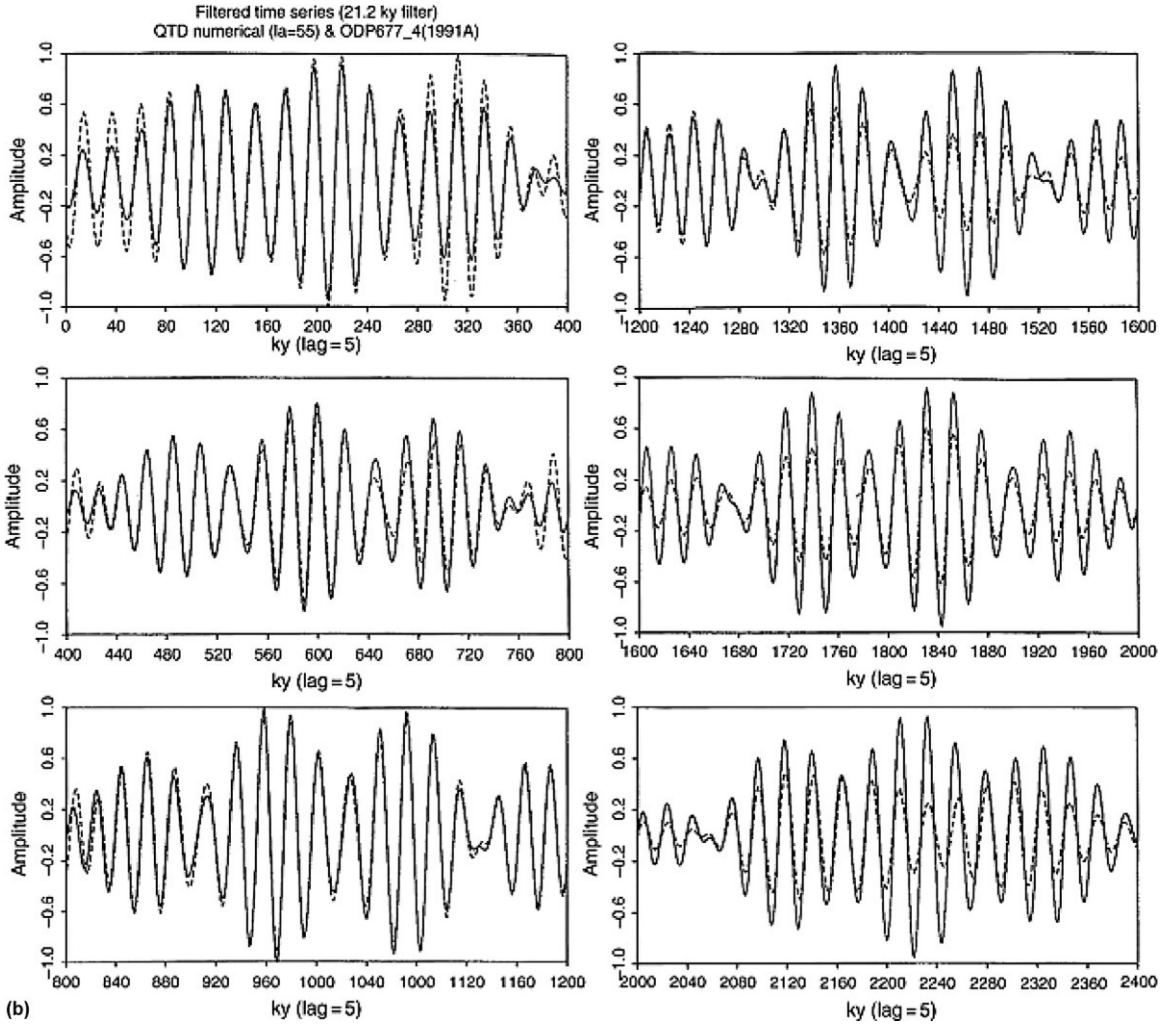


Figure 44, Cont'd (b) Same as **Figure 31(a)** but showing the coherence in the eccentricity–precession band between astronomical forcing and oxygen-isotopic (climate) response. This is the original work that led to the results reported in Shackleton et al. (1990).

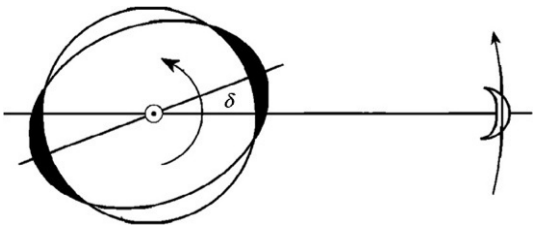


Figure 45 Schematic diagram illustrating the relationship between the tidal bulges in the oceans raised by the gravitational attraction of the Moon and the line joining the centers of mass of the Earth and Moon. The dissipation of ocean tidal energy both due to bottom friction and due to the dissipation of the internal tide leads to a phase lag of the tide and thereby to a torque that acts so as to diminish the rate of axial rotation of the planet and thereby to an increase with time of the l.o.d.

that tidal friction is considerably larger than it has been over much of the geologic past as a consequence of the fact that the ocean basins are currently of a spatial scale such that a near-resonant condition exists between the normal modes of the ocean basins and the tidal forcing insofar as the dominant M_2 tide is concerned. The interested reader will find a highly informative discussion of the tides as free oscillations of the ocean basins in Platzman (1971). This is a fascinating subject about which there is a great deal yet to be learned.

9.09.8.2 Polar Wander and the Impact of Excess Oblateness due to Mantle Convection

On the longest timescales of hundreds of millions of years on which the thermal evolution and internal structure of the Earth are governed by the mantle convection process, debate continues concerning the issue as to whether the relative latitudinal positions of the continents and the variations of their positions through time are to be ascribed solely to the action of the mantle convection-driven process of continental drift and the TPW of the axis of rotation that is expected to accompany it. If it were possible for the relative positions of the continents with respect to the axis about which the planet rotates to be spontaneously modified by a TPW event unrelated to mantle convective mixing, then it would be possible to misinterpret paleomagnetic inferences of paleolatitude as

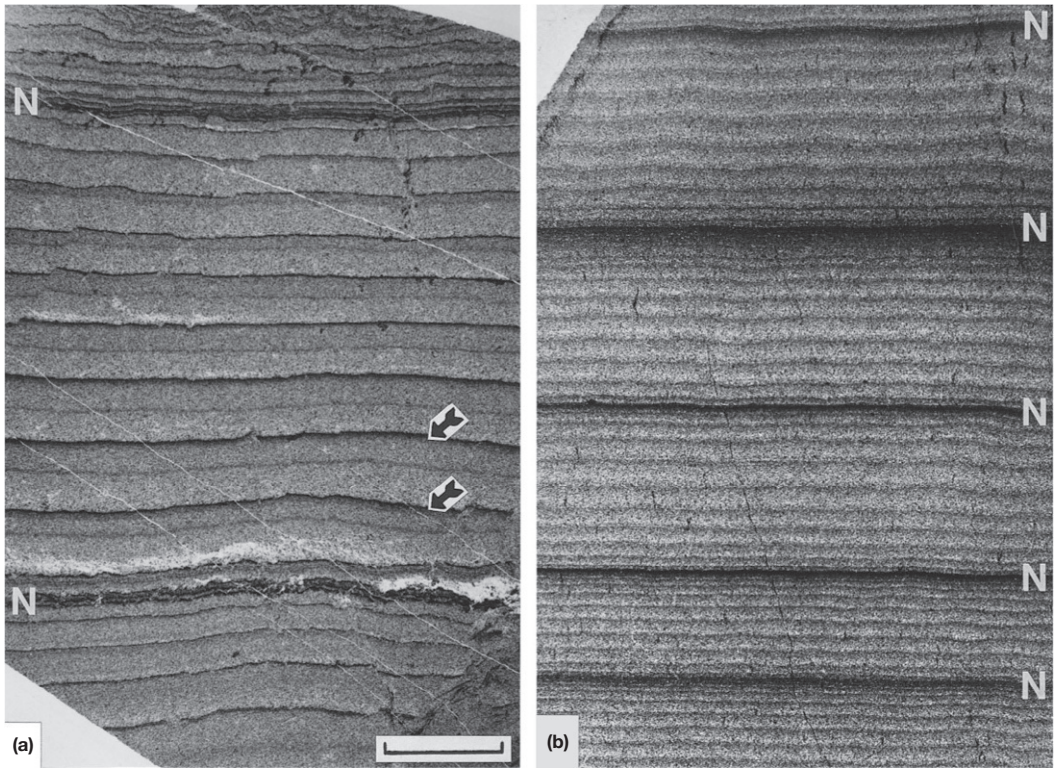


Figure 46 A typical section of tidal rhythmite from the Elatina sedimentary formation of Australia. The banding of the sediments may be interpreted in terms of the variability with time of the amplitude of the ocean tide as Earth executes a single period in its annual orbit around the Sun and the orientation of the Moon with respect to the Earth also varies simultaneously. Reproduced from Williams GE (1997) Precambrian length of day and the validity of tidal rhythmite paleotidal values. *Geophysical Research Letters* 24(4): 421–424.

arising solely due to the action of the convection process when in fact, these were due to the influence of TPW induced by some other causes. The recent resurgence of interest in TPW has come to be focused upon the possibility of occurrence of a so-called inertial interchange instability (e.g., Fisher, 1974; Goldreich and Toomre, 1969; Kirschvink et al., 1997) during which, as a consequence of the axis of greatest moment of inertia being 'suddenly' (on geologic timescales) changed, the geographic positions of the continents with respect to the rotation pole would be subject to a rapid secular shift.

On such long timescales, the issue arises once more as to whether the magnitude of the equatorial bulge represented by the fluid Love number k_f should be taken to be adequately explained by the viscoelastic field theory of Peltier (1974) in which the EQM assumption of Munk and MacDonald (1960) is invoked or whether one must allow for the existence of a bulge whose amplitude is inflated by the convection process by an amount similar to that that is apparently characteristic of the present day, as previously discussed in **Section 9.09.3.5**. Although it has been demonstrated explicitly that the impact of this influence is small insofar as the GIA prediction of polar wander speed and direction for the Late Pleistocene ice-age cycle are concerned (**Section 9.09.5.2**), on the much longer timescales associated with the mantle convection process, the impact of an equatorial bulge of amplitude larger than predicted by $k_2^T(s=0)$ in the limit of zero lithospheric thickness could be much more significant. Here, it will suffice to provide only commentary on a series of recently published or completed analyses.

In Peltier et al. (2007) and Liu and Peltier (2010, 2011), a series of analyses of the snowball Earth events that are hypothesized to have dominated the climate of the Neoproterozoic aeon (e.g., Hoffman and Schrag (2000, 2002)) have been produced. In Peltier et al. (2007), a feedback mechanism was suggested that could have prevented the occurrence of the deep glaciations that some have imagined have occurred, a mechanism that was further explored in the aforementioned Liu and Peltier (2010, 2011) contributions. More recently, further analyses have been performed in Liu and Peltier (2013a,b) to directly investigate the variations in global sea level that would have accompanied these severe glaciation events. Because the timescale of these events is such that they are expected to have lasted at least several million years, they have provided an extremely useful means of investigating the TPW that would have been induced by such a long-timescale surface loading event. For the purpose of these analyses, the version of the SLE formalism described in **Section 9.09.3.5** was employed, which enables a direct assessment of the extent to which such a surface ice-sheet loading event might be expected to produce a significant wander of the rotation pole. These analyses demonstrate that if the excess ellipticity that existed during the glaciation process was of the same order as exists at present, then the excess bulge would have been sufficient to prevent any significant TPW from occurring. Of course, we have no way of knowing whether

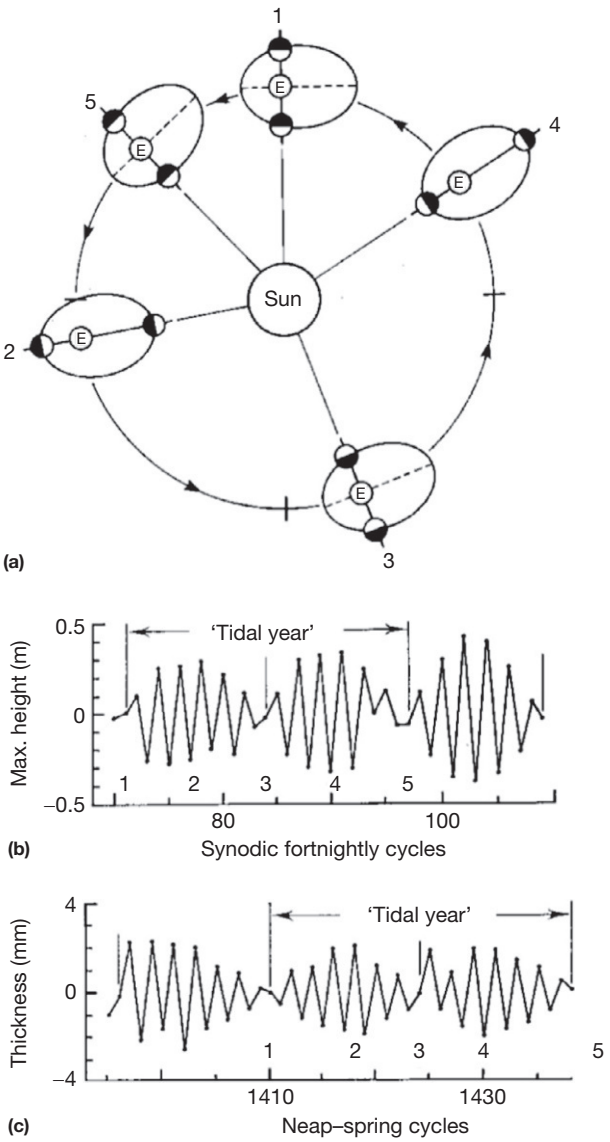


Figure 47 The manner in which the thickness of individual sedimentary bands in the rhythmite shown in **Figure 46** is interpreted in terms of the variations of tidal amplitude between 'spring' and 'neap.' Also shown is the relationship between these phases of the tide and the orbits of the Earth and Moon. Reproduced from Williams GE (2000) Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit. *Reviews of Geophysics* 38: 37–59.

the modern circumstance is at all characteristic of that that existed in this period of the distant past.

There does exist, however, one potential phenomenon that has been suggested to be a characteristic of the mantle convection process that could, at least in principle, exert 'coherent control' upon the Earth's rotation on a relatively short timescale and which is derivative of the mantle convection process itself. This is the so-called avalanche effect discussed at length by Peltier and Solheim (1992), by Solheim and Peltier (1994a,b), and more recently in Shahnas and Peltier (2010), an effect upon the temporal variability of the convection process that is controlled by the action of the phase transition from spinel to a mixture of perovskite and magnesiowüstite that exists at 660 km depth and which separates the so-called transition zone of the mantle above this level from the lower mantle beneath. The avalanche effect is found to be strong in models of the mantle convection process in which the effective viscosity of the planetary mantle is sufficiently low, as it is in the VM2 and VM5a models that have been developed on the basis of fits to observations of the GIA process discussed in the previous sections of this chapter (see **Figure 27**). Peltier (1998) had argued that this model, or a close relative derived on the basis of analyses of the GIA process, may be equally applicable on the much longer timescale characteristic of convection in the Earth's mantle, although this issue remains a subject of active debate.

Insofar as effort is concerned to confirm the possibility that IITPW may have occurred in the past, a primary era of the Earth history that has been suggested as a candidate has been the early Cambrian period (Kirschvink et al., 1997). Of course, this is an extraordinarily important period of time from the perspective of Earth evolution in general and biological evolution in particular and from the perspective of tectonophysics. From the perspective of biological evolution, it was the time of especially rapid biological diversification that occurred during the so-called Cambrian Explosion of Life. From the perspective of tectonophysics, it was a time immediately following the final phase of the Neoproterozoic breakup of the supercontinent of Rodinia during which the previously discussed series of especially severe episodes of global glaciation are believed to have occurred (e.g., Hoffman and Schrag, 2000, 2002; Kirschvink, 1992). These hypothesized glaciation events have been termed 'snowball glaciations' during which it is imagined that the surface of the Earth may have become entirely ice-covered, the oceans by a thick veneer of sea ice, and the continents by thick continental ice sheets such as those that were confined primarily to the north polar regions during times of ice-age maxima within the Pleistocene era. Although the plausibility of occurrence of the 'hard snowball' state envisioned by these authors has been questioned (e.g., Hyde et al, 2000; Liu and Peltier, 2010, 2011; Peltier et al., 2004, 2007), there is no doubt that the period of transition between the Neoproterozoic and the Cambrian eras, a transition that occurred ~541 million years ago, was a time in Earth history during which plausibly large variations in the elements of the moment of inertia tensor of the planet could have occurred. Not only was the process of mantle convection highly disturbed during the process of the rifting and dispersal of Rodinia, but also there were highly significant exchanges of mass occurring between the oceans and the highly glaciated continents at the end of the final Neoproterozoic snowball event. Li et al. (2004) had in fact proposed that the inferred equatorial location of this supercontinent during the earliest (Sturtian) episode of intense Neoproterozoic glaciation could have arisen as a consequence of a complete 90° inertial interchange instability event during which the supercontinent was shifted from the pole to the equator. It may therefore be the case that the Sturtian ice mass was not emplaced on the supercontinent while it was at the equator but rather when it was in the initial polar location. This is an idea that is suggested as a possibility by the Li et al. analysis of the paleomagnetic data but that was not explicitly envisioned by them.

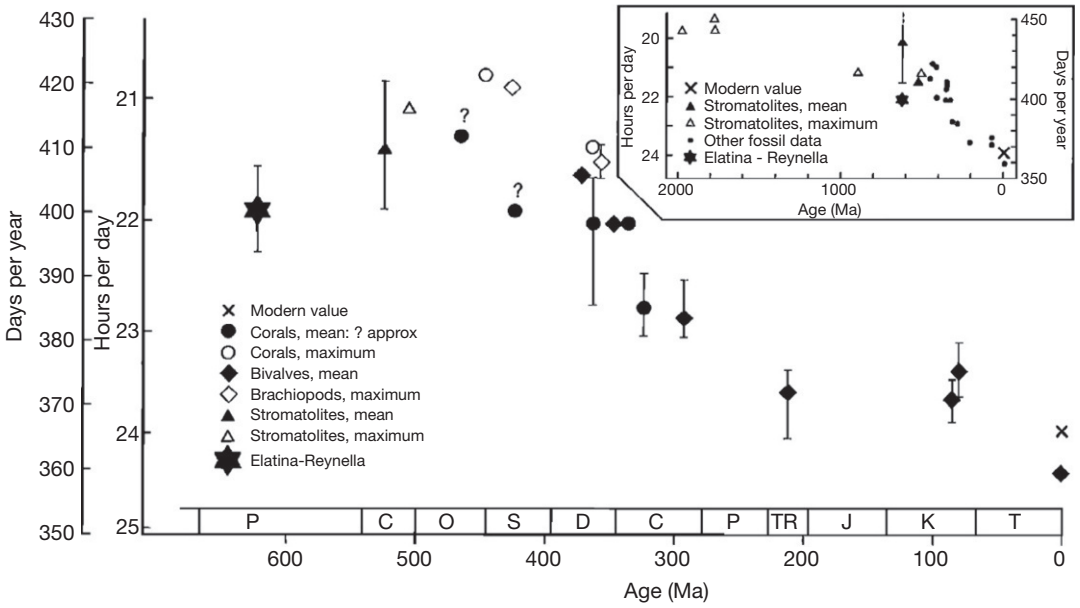


Figure 48 A relatively complete collection of estimates of the variations in the l.o.d. that have been derived from various paleontological proxy data aside from tidal rhythmites. Reproduced from Williams GE (2000) Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit. *Reviews of Geophysics* 38: 37–59.

Figure 49 illustrates the temporal interrelationships that existed in the Earth system during this critical period in Earth history between the tectonophysical events associated with the mantle convection process and the events that were simultaneously occurring in surface climate variability. The climate-related signal shown in this time series is for the carbon isotopic anomaly denoted $\delta^{13}C$, which is measured in carbonate rocks on land such as those from Namibia, which constitute the type section for much of the work that has been accomplished in support of the snowball Earth hypothesis (Hoffman and Schrag, 2000). The importance of this isotopic measurement derives from the fact that it is strongly influenced by photosynthetic activity as an isotopic fractionation occurs such that the organic matter produced in photosynthesis is enriched in ^{12}C relative to the heavier isotope ^{13}C. A time when the $\delta^{13}C$ that is measured in carbonate rocks that were precipitated from the ocean is high therefore corresponds to a time of high photosynthetic biological productivity. This is because such episodes correspond to times, as indicated on the figure, when the rate of burial of the organic matter produced in photosynthesis is significantly elevated, leaving behind an ocean that is enriched in ^{13}C. The sharp minima that are observed to characterize the time series shown in Figure 49 have therefore been taken to correspond to times in Earth history of intense 'snowball' glaciation during which photosynthetic activity was either sharply diminished or eliminated altogether as would have been the case during a 'hard snowball' event. Also indicated in Figure 49 is the timing of the episodes of intense tectonic activity that would have been associated with the breakup of the supercontinents of Rodinia and, much later during the Carboniferous period, Pangaea. The period of transition between the Neoproterozoic and the Cambrian that Kirschvink has suggested as a possible time of occurrence of an inertial interchange instability event was therefore a time of profound change in Earth history, perhaps the most important such period since the birth of the planet from the primitive solar nebula.

In an attempt to establish the early Cambrian IITPW hypothesis of Kirschvink et al. (1997) as viable, Mound et al (1999) performed a calculation of the expected impact of such an event upon relative sea-level history. Their idea was to see whether observations of relative sea-level history over timescales of millions of years might be employed as a means of determining whether an IITPW event had occurred and the analyses presented were interpreted to imply that the hypothesis not only was viable but also was supported by their initial analysis of the observational constraints they described. An important issue concerning the validity of these analyses concerns the strong dependence of the results obtained upon the thickness of the lithosphere assumed to be characteristic of the Maxwell viscoelastic model employed for the purpose of analyzing the sea-level response to the rotational excitation. Their assumption in this analysis was that the lithosphere could be treated as a complete 'unbroken-by-plate-tectonics' elastic shell over the tens of millions of years during which the IITPW event was imagined to occur. Their results, however, also suggested a very strong dependence upon the thickness assumed such that the response rapidly diminished as the assumed thickness of the lithosphere was reduced. In Sections 9.09.3 and 9.09.5 of this chapter, the question as to how the thickness of the lithosphere was to be treated, insofar as the computation of the rotational response to surface glaciation and deglaciation is concerned, was discussed in detail. There, it was suggested that appropriate treatment of the stabilizing influence of the Earth's equatorial rotational bulge required that the properties of the Maxwell viscoelastic models employed should be such that, although it would be

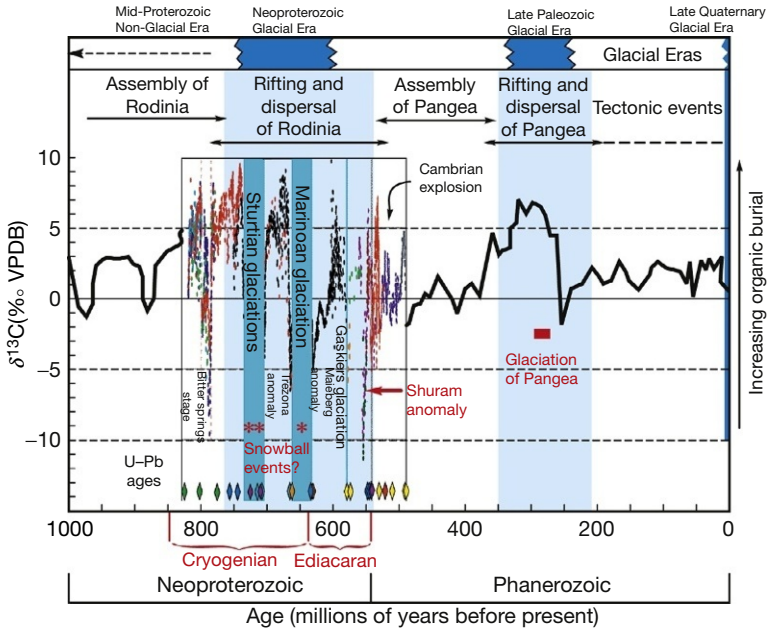


Figure 49 The variation of carbon isotopes, denoted $\delta^{13}C$, over the last 1.6 Ga of the Earth history shown together with the times of occurrence of periods of significant continental glaciation and the times of creation and destruction of supercontinents that accompany the so-called Wilson cycle. The $\delta^{13}C$ time series is from Kauffman (1997) and Hoffman and Schrag (2000), whereas the inset for the Neoproterozoic aeon is based upon the composite data set from Halverson et al. (2005). This figure was originally published in Peltier et al. (2007).

necessary to include any difference between k_f and $k_2^T(s=0)$, this difference should be that appropriate to the limit in which the thickness of the elastic lithosphere is taken to vanish in order that the calculation of the rotational response to the glaciation–deglaciation process produces physically reasonable results. As mentioned in the preceding text in commenting upon the recent results of Liu and Peltier (2013a,b), on the longest timescales, the theory of Peltier and Luthcke (2009) predicts a strong impact of any excess equatorial bulge even in the limit that this excess is taken to be that arising when the elastic thickness of the lithosphere is taken to vanish. It is interesting that the Mound et al. (1999) test of the IITPW hypothesis was not made using this form of the theory but rather a form in which the finite strength of the lithosphere was assumed to persist in the long-timescale limit. This issue should not be taken to rule out the possibility of occurrence of IITPW, however, only to question the validity of the proposed test of the hypothesis in terms of long-timescale variations of relative sea level. It may in fact be possible during a period of rapid reorganization of the convection process, when the strong correlation that exists under modern plate tectonic conditions between plate boundaries and deep mantle upwellings and downwellings may no longer obtain. In such conditions, an influence of finite lithospheric thickness may therefore have existed in which case the Mound et al. results could be applicable.

In searching for a sufficiently intense process of reorganization of the convective circulation that might be responsible for an IITPW event, such as that suggested by Kirschvink to have occurred in the Early Cambrian or Li et al. in the Neoproterozoic, I would be inclined to focus upon the times of occurrence of events related to the supercontinent cycle such as the rifting process during which an initially compact supercontinent is fragmented following the occurrence of 'avalanche' events during which the convective circulation may have made a rather sudden transition from a state of whole mantle flow to a state of layered convection (e.g., see Peltier (1996c) for a review). Peltier et al. (1997) had in fact suggested a close linkage between the so-called Wilson cycle of supercontinent creation and destruction and the continuously recurrent 'avalanche effect' that some detailed fluid mechanical models of the convection process have predicted (e.g., Solheim and Peltier, 1994a,b). Such models do apparently predict the inferred timescale of several hundred million years that appears to separate the times in Earth history during which supercontinents existed. That timescale is most visibly evident in histograms of the sample density of continental rocks of a given age as a function of their age. Such suggestions of the mechanism that could underlie these most dramatic events that may have occurred in the history of the Earth's rotation must remain a subject of speculation in search of firm quantitative justification.

References

Albero MC and Angiolini FE (1983) INGEIS Radiocarbon Laboratory dates I. *Radiocarbon* 25: 831–842.

Albero MC and Angiolini FE (1985) INGEIS Radiocarbon Laboratory dates II. *Radiocarbon* 27: 315–337.

Angulo RJ and Lessa GC (1997) The Brazilian sea level curves: A critical review with emphasis on the curves from Paranagua and Cananeia regions. *Geology* 140: 141–166.

Argus DF, Blewitt G, Peltier WR, and Kreemer C (2011) Rise of the Ellsworth Mountains and parts of the East Antarctic coast observed with GPS. *Geophysical Research Letters* 38: L16303. http://dx.doi.org/10.1029/2011GL048025,2011.

Argus DF and Gross RS (2004) An estimate of motion between the spin axis and the hotspots over the past century. *Geophysical Research Letters* 31: L06614. http://dx.doi.org/10.1029/2004GL019657.

Argus DF and Peltier WR (2010) Constraining models of postglacial rebound using space geodesy: A detailed assessment of model ICE-5G (VM2) and its relatives. *Geophysical Journal International* 181: 607–723. http://dx.doi.org/10.1111/j.1365-246X.2010.

Argus DF, Peltier WR, Drummond R, and Moore AW (2014) The Antarctic component of postglacial rebound model ICE-6G_C (VM5a) based upon GPS positioning, exposure age dating of ice thickness and relative sea level histories. *Geophysical Journal International* 198: 537–563.

Argus DF, Peltier WR, and Watkins MM (1999) Glacial isostatic adjustment observed using very long baseline interferometry and satellite laser ranging geodesy. *Journal of Geophysical Research* 104: 29077–29093.

Baksi AK (1992) $^{40}Ar/^{39}Ar$ dating of the Brunhes-Matuyama geomagnetic field reversal. *Science* 256: 356–359.

Belperio AP, Harvey N, and Bousman RP (2002) Spatial and temporal variability in the Holocene sea-level record of the South Australian coastline. *Sedimentary Geology* 150: 153–169.

Berger A (1978) Long-term variations of daily insolation and Quaternary climatic changes. *Journal of the Atmospheric Sciences* 35: 2362–2367.

Berger A, Imbrie J, Hays JD, Martinsen DG, et al. (1984) In: Berger A, et al. (eds.) The Orbit Theory of Pleistocene Climate: Support from a Revised Chronology of the Marine $\delta^{18}0$ Record in Milankovitch and Climate, pp. 269–306. Norwell, MA: D. Reidel.

Bezerra F, Barreto A, and Sugio K (2003) Holocene sea-level history on the Rio Grande do Norte State coast, Brazil. *Marine Geology* 196: 73–89.

Broecker WS and Van Donk J (1970) Insolation changes, ice volumes, and the $\delta^{18}0$ record in deep sea cores. *Reviews of Geophysics and Space Physics* 8: 169–198.

Butterworth S (1930) Theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer* 7: 536–541.

Calaya MA, Wahr JM, and Bryan FO (1999) Climate-driven polar motion. *Journal of Geophysical Research* 104: 12813–12829.

Camoin GF, Monataggioni LF, and Braithwaite CJR (1984) Late glacial to postglacial sea levels in the Western Indian Ocean. *Marine Geology* 206: 119–146.

Camoin GF, Colonna M, Montaggioni LF, Casanova J, Faure G, and Thomassin BA (1997) Holocene sea level changes and reef development in the southwestern Indian Ocean. *Coral Reefs* 16: 247–259.

Carter WE, Robertson DS, Pyle TE, and Diamante J (1986) The application of geodetic radio interferometric surveying to the monitoring of sea level. *Geophysical Journal of the Royal Astronomical Society* 87: 3–13.

Cazenave A, Dominh K, Guienhut S, et al. (2009) Sea level budget over 2003-2008: A re-evaluation from GRACE space gravimetry, satellite altimetry and Argo. *Global and Planetary Change* 65: 83–88. http://dx.doi.org/10.1016/gloplacha.2008.10.004.

Chambers DP, Wahr J, Tamisiea ME, and Nerem RS (2010) Ocean mass from GRACE and glacial isostatic adjustment. *Journal of Geophysical Research* 115: B11425. http://dx.doi.org/10.1029/2010JB007530.

Chandler SC (1891) On the variation of latitude, I. *Astronomical Journal* 11: 59–61.

Chao BF and Gross RS (1987) Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes. *Geophysical Journal of the Royal Astronomical Society* 91: 569–596.

Chao BF, O'Conner WP, Chang ATC, Hal DK, and Foster JL (1987) Snow load effect on the Earth's rotation and gravitational field, 1979-1985. *Journal of Geophysical Research* 92: 9415–9422.

Cheng MK, Eanes RJ, Shum CK, Schutz BE, and Tapley BD (1989) Temporal variations in low degree zonal harmonics from Starlette orbit analysis. *Geophysical Research Letters* 16: 393–396.

Cheng M, Ries JC, and Tapley BD (2011) Variations in Earth's figure axis from satellite Laser ranging and GRACE. *Journal of Geophysical Research* 116: B01409. http://dx.doi.org/10.1029/2010JB000850.

Cheng M and Tapley BD (2004) Variations in the earth's oblateness during the past 28 years. *Journal of Geophysical Research* 109: B09402. http://dx.doi.org/10.1029/2004JB003028.

Church JA, White NJ, Konikow LF, et al. (2011) Revisiting the earth's sea-level and energy budgets from 1961-2008. *Geophysical Research Letters* 38: L18601. http://dx.doi.org/10.1029/2011GL048794.

Clark JA (1976) Greenland's rapid postglacial emergence: A result of ice-water gravitational attraction. *Geology* 4: 310–312.

Clark JA, Farrell WE, and Peltier WR (1978) Global changes in postglacial sea levels: A numerical calculation. *Quaternary Research* 9: 265–287.

Clark JA and Lingle CS (1977) Future sea level changes due to West Antarctic ice-sheet fluctuations. *Nature* 269: 206–209.

Clark PU, Mitrovica JX, Milne GA, and Tamasea ME (2002) Sea level finger printing as a direct test of the source of meltwater pulse 1a. *Science* 295: 2438–2441.

Codignotto JD, Kokott RR, and Marcomini SC (1992) Neotectonism and sea-level changes in the coastal zone of Argentina. *Journal of Coastal Research* 8: 125–133.

Cox CM and Chao BF (2002) Detection of a large scale mass redistribution in the terrestrial system since 1998. *Science* 297: 831–833.

Dahlen FA (1976) The passive influence of the oceans upon the rotation of the Earth. *Geophysical Journal of the Royal Astronomical Society* 46: 406–663.

Dahlen FA and Smith ML (1975) The influence of rotation on the free oscillations of the Earth. *Philosophical Transactions of the Royal Society of London, Series A* 267: 583–629.

Dalmau W (1997) Critical remarks on the use of medieval eclipse records for the determination of long-term changes in the Earth's rotation. *Surveys in Geophysics* 18: 213–223.

Daly L and Le Goff M (1996) An updated and homogeneous world secular variation database. Part 1. Smoothing of the archaeomagnetic results. *Physics of the Earth and Planetary Interiors* 93: 159–190.

Deblonde G and Peltier WR (1991) A one dimensional model of continental ice-volume fluctuations through the Pleistocene: Implications for the origin of the mid-Pleistocene climate transition. *Journal of Climate* 4: 18–34.

Deblonde G and Peltier WR (1993) Late Pleistocene ice-age scenarios based upon observational evidence. *Journal of Climate* 6: 709–727.

Delibrias G and Laborel J (1971) Recent variations of the sea level along the Brazilian coast. *Quaternia* 14: 45–49.

Dickey JO, Marcus SL, de Viron O, and Fukumori I (2002) Recent Earth oblateness variations: Unravelling climate and postglacial rebound effects. *Science* 298: 1975–1977.

Dickman SR (1977) Secular trend of the Earth's rotation pole: Consideration of motion of the latitude observatories. *Geophysical Journal of the Royal Astronomical Society* 57: 41–50.

Domack E, Duran D, Leventer A, et al. (2005) Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch. *Nature* 436: 681–685.

Dunberry M and Bloxham J (2006) Azimuthal flows in the Earth's core and changes in length of day at millennial timescales. *Geophysical Journal International* 165: 32–46.

Dyurgerov MB (2010) Reanalysis of Glacier Changes from the IGY to the IPY, 1960–2008. Publication 108. Data on Glaciological Studies, 116 pp, Moscow.

Dyurgerov MB and Meier MF (2005) Glaciers and the changing Earth system: A 2004 snapshot. *Occassional Paper No. 58*. Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Engelhart SE, Horton BP, Douglas BC, Peltier WR, and Tornqvist TE (2010) Spatial variability of Late Holocene and 20th century sea-level rise along the Atlantic coast of the United States. *Geology* 37: 1115–1118. http://dx.doi.org/10.1130/G30360A.1.

Engelhart SE, Peltier WR, and Horton BP (2011) Holocene relative sea level changes and glacial isostatic adjustment of the U.S. Atlantic coast. *Geology* 39: 751–754. http://dx.doi.org/10.1030/G31857.1.

Fairbanks RG (1989) A 17,000 year glacio-eustatic sea level record:influence of glacial melting rates on the Younger Dryas event and deep ocean circulation. *Nature* 342: 637–642. http://dx.doi.org/10.1038/342637a0.

Fairbridge RG (1976) Shellfish-eating preceramic indians in coastal Brazil. *Science* 191: 353–359.

Farrell WE (1972) Deformation of the Earth by surface loads. *Reviews of Geophysics* 10: 761–797.

Farrell WE and Clark JA (1976) On postglacial sea level. *Geophysical Journal of the Royal Astronomical Society* 46: 647–667.

Faure H, Fontes JC, Hebrasd L, Monteillet J, and Pirazolli PA (1980) Geoidal changes and shore-level tilt along Holocene estuaries: Senegal River area, West Africa. *Science* 210: 421–423.

Faure H and Hebrard L (1977) Variations des lignes de ravages au Senegalet au Mauritania au cours de l'Holocene. *Studia Geologica Polonica* 52: 143–157.

Fisher D (1974) Some more remarks on polar wandering. *Journal of Geophysical Research* 79: 4041–4045.

Gegout P and Cazanave A (1991) Geodynamic parameters derived from 7 years of laser data on LAGEOS. *Geophysical Research Letters* 18: 1739–1742.

Gibb JG (1986) A New Zealand regional Holocene eustatic sea-level curve and its application to determination of vertical tectonic movements. *Royal Society of New Zealand Bulletin* 24: 377–395.

Gold T (1955) Instability of the Earth's axis of rotation. *Nature* 175: 526–529.

Goldreich P and Toomre A (1969) Some remarks on polar wandering. *Journal of Geophysical Research* 74: 2555–2567.

Griffiths SD and Peltier WR (2008) Mega-tides in the glacial ocean and rapid climate change. *Geophysical Research Letters* 35: L08605 http://dx.doi.org/10.10292008GL033263.

Griffiths SD and Peltier WR (2009) Modeling of polar ocean tides at the Last Glacial Maximum: Amplification, sensitivity, and climatological implications. *Journal of Climate* 22(11): 2905–2924. http://dx.doi.org/10.1175/2008JCLI2540.1.

Gross RS (1986) The influence of earthquakes on the Chandler wobble during 1977-1983. *Geophysical Journal of the Royal Astronomical Society* 85: 161–177.

Gross RS (2000) The excitation of the Chandler wobble. *Geophysical Research Letters* 27: 2329–2342.

Gross R and Poutanen M (2009) Geodetic observations of glacial isostatic adjustment: Understanding glacial isostatic adjustment: A joint DynaQlim/GGOS Workshop, Espo, Finland, 22–26 June 2009. *EOS* 90(41): 356.

Gross RS and Vondrak J (1999) Astrometric and space-geodetic observations of polar wander. *Geophysical Research Letters* 26: 2085–2088. http://dx.doi.org/10.1029/1999GL900422.

Hall GF, Hill DF, Horton BP, Engelhart SE, and Peltier WR (2013) A high resolution study of tides in Delaware Bay: Past conditions and future scenarios. *Geophysical Research Letters* 40: 338–342. http://dx.doi.org/10.1029/2012GL054675.

Halverson GP, Hoffman PF, Schrag DP, Maloof AC, Rice A, and Hugh N (2005) Toward a Neoproterozoic composite carbon-isotopic record. *GSA Bulletin* 117: 1181–1207.

Hays JD, Imbrie J, and Shackleton NJ (1976) Variations in the earth's orbit: Pacemaker of the ice-ages. *Science* 194: 1121–1132.

Hide R, Birch NT, Morrison LV, Shea DJ, and White AA (1980) Atmospheric angular momentum fluctuations an changes in the length of the day. *Nature* 286: 114–117.

Hill D, Griffiths S, Peltier WR, Horton BP, and Tornquist TE (2011) High resolution numerical Modeling of tides in the Western Atlantic, Gulf of Mexico and Caribbean during the Holocene. *Journal of Geophysical Research: Oceans* 116: C10014. http://dx.doi.org/10.1029/2010JC006896.

Hoffman PF and Schrag DF (2000) Snowball Earth. *Scientific American* 282: 68–75.

Hoffman PF and Schrag DA (2002) The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* 14: 129–155.

Hongre L, Hulot G, and Khoklov A (1998) An analysis of the geomagnetic field over the past 2000 years. *Physics of the Earth and Planetary Interiors* 106: 311–335.

Horton BP, Peltier WR, Culver SJ, et al. (2009) Holocene sea level changes along the North Carolina coastline and their implications for glacial isostatic adjustment models. *Quaternary Science Reviews* 28: 1725–1736. http://dx.doi.org/10.1016/j.quascirev.2009.02.002.

Hyde WT, Crowley TJ, Baum SK, and Peltier WR (2000) Neoproterozoic "snowball Earth" simulations with a coupled climate/ice sheet model. *Nature* 405: 425–430.

Jansen E, Overpeck J, Briffa KR, et al. (2007) Palaeoclimate. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, and Miller HL (eds.) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 433–497. Cambridge: Cambridge University Press.

Jault D and Le Mouel J-L (1993) Circulation in the liquid core and coupling with the mantle. In: Singh RP, Feissel M, Tapley BD, and Shum CK (eds.) *Observations of Earth from Space. Advances in Space Research*, vol. 13. Oxford: Pergamon (11)221–233.

Kauffman AJ (1997) An ice age in the tropic. *Nature* 386: 227–228.

Kirschvink JL (1992) Late Proterozoic low-latitude global glaciation: The snowball Earth. In: Schopf JW and Klein C (eds.) *The Proterozoic Biosphere, a Multi-disciplinary Study*, pp. 51–52. Cambridge: Cambridge University Press.

Kirschvink JL, Ripperdan RL, and Evans DA (1997) Evidence for a large-scale reorganization of Early continental masses by inertial interchange true polar wander. *Science* 277: 541–545.

Koba M, Nakata T, and Takahashi T (1982) Late Holocene eustatic changes deduced from geomorphological features and their C-14 dates in the Ryukyu Islands, Japan. *Palaeogeography, Palaeoclimatology, Palaeoecology* 39: 231–260.

Korte M and Constable C (2003) Continuous global geomagnetic field models for the past 3000 years. *Physics of the Earth and Planetary Interiors* 140: 73–89.

Krabil W, Abdalati W, Frederick E, et al. (2000) Greenland Ice Sheet: High elevation balance and peripheral thinning. *Science* 289: 428–430.

Kuehne J and Wilson CR (1991) Terrestrial water storage and polar motion. *Journal of Geophysical Research* 96: 4337–4345.

Lambeck K (1980a) Changes in length-of-day and atmospheric circulation. *Nature* 286: 104–105.

Lambeck K (1980b) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. New York: Cambridge University Press, 449 pp.

Lambeck K (2002) Sea level change from mid-Holocene to recent time: An Australian example with global implications. In: Mitrovica JX and Vermeersen LA (eds.) *Ice Sheets, Sea Level and the Dynamics of the Earth*, *AGU Monographs*, vol. 29, pp. 33–50.

Lambeck K and Chappell J (2001) Sea level change through the last glacial cycle. *Science* 292(5517): 679–686.

Laskar J (1988) Secular evolution of the solar system over 10 million years. *Astronomy and Astrophysics* 198: 341–362.

Li ZX, Evans DAD, and Zhang S (2004) A 90 degree spin on Rodinia: possible causal links between the Neoproterozoic supercontinent, superplume, true polar wander and low latitude glaciation. *Earth and Planetary Science Letters* 220: 409–421.

Liu Y and Peltier WR (2010) A carbon cycle coupled model of Neoproterozoic glaciation. Part I: Influence of continental configuration on the formation of a "soft snowball" *Journal of Geophysical Research: Atmospheres* 115: D17111. http://dx.doi.org/10.1029/2009JD013082.

Liu Y and Peltier WR (2011) A carbon cycle coupled model of Neoproterozoic glaciation. Explicit carbon cycle with stochastic perturbations. *Journal of Geophysical Research: Atmospheres* 116: D02125. http://dx.doi.org/10.1029/2010JD015128.

Liu Y and Peltier WR (2013a) Sea level variations during snowball Earth formation (Part 1): Preliminary analysis. *Journal of Geophysical Research*, *Solid Earth* 118: 4410–4424. http://dx.doi.org/10.1002/jgrb.50293,2013.

Liu Y and Peltier WR (2013b) Sea level variations during snowball earth formation (Part 2): The influence of rotational feedback. *Journal of Geophysical Research: Solid Earth* 118: 4425–4445. http://dx.doi.org/10.1002/jgrb.50294,2013.

Liu Y, Peltier WR, Yang J, and Vettoretti G (2013) The initiation of Neoproterozoic "snowball" climates in CCSM3: The influence of paleocontinental configuration. *Climate of the Past* 9: 2555–2577.

MacKintosh A, Golledge N, Domack E, et al. (2011) Retreat of the Antarctic ice sheet during the last glacial termination. *Nature Geoscience* 4: 195–2002. http://dx.doi.org/10.1038/ngeo1061.

Mackintosh AN, Verleyen E, O'Brien PE, et al. (2013) Retreat history of the East Antarctic ice sheet since Last Glacial Maximum. *Quaternary Science Reviews* 33: 1–21.

Markowitz W (1960) Latitude and longitude and the secular motion of the pole. *Methods and Techniques in Geophysics*, vol. 1, 325–361.

Martin L, Sugio K, Flexor JM, Dominguez JML, and Bittencourt ACSP (1987) Quaternary evolution of the central part of the Brazilian coast, the role of relative sea-level variation and of shoreline drift. In: *Quaternary Coastal Geology of West Africa and South America. UNESCO Reports in Marine Science 43,* pp. 97–145.

Mathews J and Walker RL (1987) *Mathematical Methods of Physics*, 2nd edn. New York: Benjamin.

Mikhailov AA (1971) On the motion of the Earth's poles. *Astronomicheskii Zhurnal* 48: 1301–1304.

Milankovitch M (1940) Canon of Insolation and the Ice Age Problem. Royal Serbian Academy, Special Publications, vol. 132, Section on Math and Natural Sciences, vol. 33, Belgrade (in German, translated by the Israeli Translation Services).

Milne GA and Mitrovica JX (1996) Post-glacial sea level change on a rotating Earth: First results from a gravitationally self-consistent sea-level equation. *Geophysical Journal International* 126: F13–F20.

Mitrovica JX and Peltier WR (1991) On postglacial sea level over the equatorial ocean. *Journal of Geophysical Research* 96: 20053–20071.

Mitrovica JX, Wahr J, Matsuyama I, and Paulson A (2005) The rotational stability of an ice-age Earth. *Geophysical Journal International* 161: 491–506.

Morner NA (1991) Holocene sea level changes in the Tierra del Fuego region. *Boletim IG-USP. Publicação Especial No. 8*, Sao Paulo.

Morrison LV (1973) Rotation of the Earth and the constancy of G. *Nature* 241: 519–520.

Morrison LV and Stephenson FR (2001) Historical eclipses and the variability of the Earth's rotation. *Journal of Geodynamics* 32: 247–265.

Mound JE, Mitrovica JX, Evans DAD, and Kirschvink JL (1999) A sea-level test for inertial interchange true polar wander events. *Geophysical Journal International* 136: F5–F10.

Muller RA and Stephenson FR (1975) The acceleration of the Earth and Moon from early observations. In: Rosenburg GD and Runcorn SK (eds.) *Growth Rhythms and History of the Earth's Rotation*, pp. 459–534. New York: John Wiley.

Munk WH and MacDonald GJF (1960) *The Rotation of the Earth*. New York: Cambridge University Press.

Nakiboglu SM (1982) Hydrostatic theory of the Earth and its mechanical implications. *Physics of the Earth and Planetary Interiors* 28: 302–311.

Nakiboglu SM and Lambeck K (1980) Deglaciation effects on the rotation of the Earth. *Geophysical Journal of the Royal Astronomical Society* 62: 49–58.

Nakiboglu SM and Lambeck K (1981) Deglaciation related features of the Earth's gravity field. *Tectonophysics* 72: 289–303.

Nerem RS and Wahr J (2011) Recent changes in the Earth's oblateness driven by Greenland and Antarctica Mass loss. *Geophysical Research Letters* 38(13): L13501. http://dx.doi.org/10.1029/2011GL047879.

Newton RR (1972) *Medieval Chronicles and the Rotation of the Earth*. Baltimore: Johns Hopkins University Press.

Oba T and Irino T (2012) Sea level at the Last Glacial Maximum, constrained by oxygen isotopic curves of planktonic foraminifera in the Japan Sea. *Journal of Quaternary Science* 27(9): 941–947. http://dx.doi.org/10.1002/jqs.2585.

Peltier WR (1974) The impulse response of a Maxwell Earth. *Reviews of Geophysics and Space Physics* 12: 649–669.

Peltier WR (1976) Glacial isostatic adjustment, II: The inverse problem. *Geophysical Journal of the Royal Astronomical Society* 46: 669–706.

Peltier WR (1982) Dynamics of the ice-age Earth. *Advances in Geophysics* 24: 1–146.

Peltier WR (1983) Constraint on deep mantle viscosity from LAGEOS acceleration data. *Nature* 304: 434–436.

Peltier WR (1985) The LAGEOS constraint on deep mantle viscosity: Results from a new normal mode method for the inversion of viscoelastic relaxation spectra. *Journal of Geophysical Research* 90: 9411–9421.

Peltier WR (1994) Ice-age paleotopography. *Science* 265: 195–201.

Peltier WR (1996a) Physics of the ice-age cycle. In: Duplessy J-C (ed.) *Long Term Climate Variations*, pp. 453–481. Berlin: Springer-Verlag Inc.

Peltier WR (1996b) Mantle viscosity and ice-age ice-sheet topography. *Science* 273: 1359–1364.

Peltier WR (1996c) Phase transition modulated mixing in the mantle of the Earth. *Philosophical Transactions of the Royal Society A* 354: 1425–1447.

Peltier WR (1998) Postglacial variations in the level of the sea: Implications for climate dynamics and solid-Earth geophysics. *Reviews of Geophysics* 36: 603–689.

Peltier WR (1999) Global sea level rise and glacial isostatic adjustment. *Global and Planetary Change* 20: 93–123.

Peltier WR (2002a) Comments on the paper of Yokoyama et al. (2000) entitled "Timing of the last glacial maximum from observed sea level minima" *Quaternary Science Reviews* 21: 409–414.

Peltier WR (2002b) One eustatic sea level history, Last Glacial Maximum to Holocene. *Quaternary Science Reviews* 21: 377–396.

Peltier WR (2002c) Global glacial isostatic adjustment: paleogeodetic and space-geodetic tests of the ICE-4G (VM2) model. *Journal of Quaternary Science* 17: 491–510.

Peltier WR (2004) Global glacial isostasy and the surface of the ice-age Earth: The ICE-5G(VM2) model and GRACE. *Annual Review of Earth and Planetary Sciences* 32: 111–149.

Peltier WR (2005) On the hemispheric origins of meltwater pulse 1a. *Quaternary Science Reviews* 24: 1655–1671.

Peltier WR (2007a) Postglacial coastal evolution: Ice-Earth-Ocean interactions in a period of rapid climate change. In: Harff J, Hay WW, and Tetzlaff DM (eds.) *Coastline Changes: Interrelation of Climate and Geological Processes*. Boulder, CO: The Geological Society of America Inc, Geological Society of America Special Paper 426.

Peltier WR (2007b) History of Earth rotation. In: Schubert G (ed.) *Volume 9 of "The Treatise on Geophysics"*, pp. 243–293. Oxford, UK: Elsevier Press.

Peltier WR (2009) Closure of the budget of global sea level rise over the GRACE era: The importance and magnitudes of the required corrections for global glacial isostatic adjustment. *Quaternary Science Reviews* http://dx.doi.org/10.1016/j.quascirev.2009.04.004.

Peltier WR and Andrews JT (1976) Glacial isostatic adjustment I: the forward problem. *Geophysical Journal of the Royal Astronomical Society* 46: 605–646.

Peltier WR, Argus DF, and Drummond R (2014) Space geodesy constrains Ice-Age terminal deglaciation: The ICE-6G (VM5a) model. *Journal of Geophysical Research: Solid Earth*.

Peltier WR, Butler S, and Solheim LP (1997) The influence of phase transformations on mantle mixing and plate tectonics. In: Crossley DJ (ed.) *Earth's Deep Interior*, pp. 405–430. Amsterdam: Gordon and Breach Publishers Inc.

Peltier WR and Drummond R (2010) Deepest mantle viscosity: Constraints from earth rotation anomalies. *Geophysical Research Letters* 37: L12304. http://dx.doi.org/10.1029/2010GL043219.

Peltier WR, Drummond R, and Roy K (2012) Comment on "Ocean mass from GRACE and glacial isostatic adjustment" by D.P. Chambers et al. *Journal of Geophysical Research, Solid Earth* 117: B11403. http://dx.doi.org/10.1029/2011JB008967.

Peltier WR and Fairbanks RG (2006) Global glacial ice-volume and Last Glacial Maximum duration from an extended Barbados Sea Level record. *Quaternary Science Reviews* 25: 3322–3337.

Peltier WR, Farrell WE, and Clark JA (1978) Glacial isostasy and relative sea-level: A global finite element model. *Tectonophysics* 50: 81–110.

Peltier WR and Jiang X (1996) Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *Journal of Geophysical Research* 101: 3269–3290 (Correction, 1997, *Journal of Geophysical Research* 102: 10101–10103).

Peltier WR, Liu Y, and Crowley JW (2007) Snowball Earth prevention by dissolved organic carbon remineralization. *Nature* 450: 813–818, http://dx.doi.org/10.1038.Nature06354,2007.

Peltier WR and Luthcke SB (2009) On the origins of earth rotation anomalies: New insights on the basis of both "paleogeodetic" data and Gravity Recovery and Climate Experiment (GRACE) data. *Journal of Geophysical Research: Solid Earth* 114: B11405. http://dx.doi.org/10.1029/2009JB006352.2009.

Peltier WR, Shennan I, Drummond R, and Horton B (2002) On the postglacial isostatic adjustment of the British Isles and the shallow viscoelastic structure of the Earth. *Geophysical Journal International* 148: 443–475.

Peltier WR and Solheim LP (1992) Mantle phase transitions and layered chaotic convection. *Geophysical Research Letters* 19: 321–324.

Peltier WR and Solheim LP (2004) The climate of the Earth at Last Glacial Maximum: Statistical equilibrium state and a mode of internal variability. *Quaternary Science Reviews* 23: 335–357.

Peltier WR, Tarasov L, Vettoretti G, and Solheim LP (2004) Climate dynamics in deep time: Modeling the "snowball bifurcation" and assessing the plausibility of its occurrence. In: Jenkins G, McMennamin MAS, McKay CP, and Sohl L (eds.) *The Extreme Proterozoic: Geology, Geochemistry, and Climate. AGU Monograph*, vol. 146, pp. 107–124. Washington, DC: AGU Press.

Pirazzoli PA (1978) High stands of Holocene sea level in Northwest Pacific. *Quaternary Research* 10(1): 1–29.

Platzman GW (1971) Ocean tides and related waves. In: Reid WH (ed.) *Mathematical Problems in the Geophysical Sciences. Inverse Problems, Dynamo Theory and Tides*, vol. 2. Providence Rhode Island: American Mathematical Society.

Ponte RM, Stammer D, and Marshall J (1998) Oceanic signals in observed motions of the Earth's pole of rotation. *Nature* 391: 476–479.

Porter SC, Stuiver M, and Heusser CJ (1984) Holocene sea level changes along the Strait of Magellan and Beagle Channel, Southernmost South America. *Quaternary Research* 22(1): 59–67.

Proverbio E and Queseda V (1974) Secular variation in latitudes and longitudes and continental drift. *Journal of Geophysical Research* 79: 4941–4943.

Quinn TR, Tremaine S, and Duncan M (1991) A three million year integration of the Earth's orbit. *Astronomical Journal* 101: 2287–2305.

Ratcliff JT and Gross RS (2010) Combinations of Earth orientation measurements: SPACE2009, COMB2009 and POLE2009. Jet propulsion Laboratory Pbl. 10-22, 29 pp., Pasedena, CA.

Rochester MG and Smiley DE (1965) Geomagnetic core-mantle coupling and the Chandler wobble. *Geophysical Journal of the Royal Astronomical Society* 10: 289–315.

Rodell M, Houser PR, Jambor U, et al. (2004) The global Ly Greenland and data assimilation system. Bull. *American Meteorological Society* 85: 381–391.

Rostami K, Peltier WR, and Mangini A (2000) Quaternary marine terraces, sea level changes and uplift history of Patagonia, Argentina: Comparisons with predictions of the ICE-4G (VM2) model of the global process of glacial isostatic adjustment. *Quaternary Science Reviews* 19: 1495–1525.

Roy K and Peltier WR (2011) GRACE era secular trends in Earth rotation parameters: A global scale impact of the global warming process? *Geophysical Research Letters* 38: L10306. http://dx.doi.org/10.1029/2011GL047282.2011.

Rubincam DR (1984) Postglacial rebound observed by LAGEOS and the effective viscosity of the lower mantle. *Journal of Geophysical Research* 89: 1077–1087.

Sabadini R and Peltier WR (1981) Pleistocene deglaciation and the Earth's rotation: implications for mantle viscosity. *Geophysical Journal of the Royal Astronomical Society* 66: 553–578.

Salehipour H, Stuhne GD, and Peltier WR (2013) A higher order discontinuous Galerkin, global shallow water model: global ocean tides and aquaplanet benchmarks. *Ocean Modelling* 69: 93–107.

Sauriau A and Cazanave A (1985) Re-evaluation of the seismic excitation of the Chandler wobble from recent data. *Earth and Planetary Science Letters* 75: 410–416.

Schofield JC (1975) Sea-level fluctuations cause periodic post-glacial progradation, South Kaipara Barrier, North Island, New Zealand. *New Zealand Journal of Geology and Geophysics* 18: 295–316.

Shackleton NJ (1967) Oxygen isotope analysis and Pleistocene temperatures re-addressed. *Nature* 215: 15–17.

Shackleton NJ (2000) The 100,000 year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* 289: 1897–1902.

Shackleton NJ, Berger A, and Peltier WR (1990) An alternative astronomical calibration of the lower Pleistocene timescale based upon ODP 677. *Transactions of the Royal Society of Edinburgh: Earth Sciences* 81: 251–261.

Shackleton NJ and Opdyke ND (1973) Oxygen isotope and paleomagnetic stratigraphy of equatorial Pacific core V28-238: Oxygen isotope temperatures and ice volumes on a 10^5-year time scale. *Quaternary Research* 3: 39–55.

Shahnas H and Peltier WR (2010) Layered convection and the impacts of the Perovskite-post Perovskite phase transition on mantle dynamics under isochemical

conditions. *Journal of Geophysical Research: Solid Earth* 115: B11408. http://dx.doi.org/10.1029/2009JB007199.
Solheim LP and Peltier WR (1994a) Avalanche effects in phase transition modulated thermal convection: A model of earth's mantle. *Journal of Geophysical Research* 99: 6997–7018.
Solheim LP and Peltier WR (1994b) 660 km phase boundary deflections and episodically layered isochemical convection. *Journal of Geophysical Research* 99: 15861–15875.
Stephenson ER and Morrison LV (1995) Long term fluctuations in the earth's rotation: 700 B.C. to A.D. 1990. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 351: 165–202.
Stuiver M and Reimer PJ (1993) Extended ^{14}C data base and revised calib. 3.0 ^{14}C age calibration program. *Radiocarbon* 35: 215–230.
Sugihara K, Nakamori T, Iryu Y, Sasakai K, and Blanchon P (2003) Holocene sea-level change and tectonic uplift deduced from raised reef terraces, Kikai-jima, Ryukyu Islands, Japan. *Sedimentary Geology* 159: 5–25.
Tapley BD, Bettadpur JC, Ries JC, Thompson PF, and Watkins M (2004) GRACE measurements of mass variability in the Earth system. *Science* 305: 503–505.
Tarantola A and Valette B (1982a) Inverse problems = quest for information. *Journal of Geophysics* 50: 159–170.
Tarantola A and Valette B (1982b) Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics* 20: 219–232.
Tarasov L and Peltier WR (2002) Greenland glacial history and local geodynamic consequences. *Geophysical Journal International* 150: 198–229.
Tarasov L and Peltier WR (2005) Arctic freshwater forcing of the Younger-Dryas cold reversal. *Nature* 435: 662–665.
Tarasov L and Peltier WR (2006) A calibrated deglacial chronology for the North American continent: Evidence of an Arctic trigger for the Younger-Dryas event. *Quaternary Science Reviews* 25: 659–688.
Thomas R, Frederik E, Krabill W, Manizade S, and Martin C (2006) Progressive increase in ice loss from Greenland. *Geophysical Research Letters* 33: L10503. http://dx.doi.org/10.1029/2006GL026075.
Valastro S Jr., Davis EM, Vallrela AG, and Ekland-Olson C (1980) University of Texas at Austin radiocarbon dates XIV. *Radiocarbon* 22: 1090–1115.
Velicogna I and Wahr J (2006a) Acceleration of Greenland ice mass loss in spring 2004. *Nature* 443(7109): 329–331.
Velicogna I and Wahr J (2006b) Measurements of time variable gravity show mass loss in Antarctica. *Science* 311: 1754–1756.
Vincente RO and Yumi S (1969) Co-ordinates of the pole (1899-1968) returned to the conventional international origin. *Publications of the International Latitude Observatory of Mizusawa* 7: 41–50.
Vincente RO and Yumi S (1970) Revised values (1941-1961) of the co-ordinates of the pole referred to the CIO. *Publications of the International Latitude Observatory of Mizusawa* 7: 109–112.
Vondrak J (1984) Long period behaviour of polar motion between 1900.0 and 1980.0. *Annals of Geophysics* 3: 351–356.
Waelbroeck C, Labyrie L, Michel E, et al. (2002) Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quaternary Science Reviews* 21(1–3): 295–305.
Wahr JM (1982) The effects of the atmosphere and oceans on the Earth's wobble-I. Theory. *Geophysical Journal of the Royal Astronomical Society* 70: 349–372.
Wahr JM (1983) The effects of the atmosphere and oceans on the Earth's wobble and on the seasonal variations in the length of day-II. Results. *Geophysical Journal of the Royal Astronomical Society* 74: 451–487.
Widder DV (1946) *The Laplace Transform*. Princeton, NJ: Princeton University Press.
Williams GE (1997) Precambrian length of day and the validity of tidal rhythmite paleotidal values. *Geophysical Research Letters* 24(4): 421–424.
Williams GE (1998) Precambrian tidal and glacial clastic deposits: Implications for Precambrian Earth-Moon dynamics and paleoclimate. *Sedimentary Geology* 120: 53–74.
Williams GE (2000) Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit. *Reviews of Geophysics* 38: 37–59.
Wilson CR and Haubrich RA (1976) Meteorological excitation of the Earth's wobble. *Geophysical Journal of the Royal Astronomical Society* 46: 707–743.
Wilson CR and Vincente RO (1990) Maximum likelihood estimates of polar motion parameters. In: McCarthy DD and Carter WE (eds.) *Variations in Earth Rotation*. *American Geophysical Union Monograph Series*, vol 59, pp. 151–155. Washington, DC: AGU.
Wu P and Peltier WR (1984) Pleistocene deglaciation and the Earth's rotation: A new analysis. *Geophysical Journal of the Royal Astronomical Society* 76: 202–242.
Yoder CF (1995) www.agu.org/references/geophys/4_Yoder.pdf.
Yoder CF, Williams JG, Dickey JO, Schutz BE, Eanes RJ, and Tapley BD (1983) Secular variation of the Earth's gravitational harmonic J_2 coefficient from LAGEOS and nontidal acceleration of Earth rotation. *Nature* 303: 757–762.
Yokoyama Y, Lambeck K, DeDekkar P, Johnston P, and Fifield LK (2000) Timing of the Last Glacial Maximum from observed sea level minima. *Nature* 406: 713–716 (Correction, 2001, *Nature* 412: 99).

This page intentionally left blank

9.10 Coevolution of Life and Earth

GJ Retallack, University of Oregon, Eugene, OR, USA

© 2015 Elsevier B.V. All rights reserved.

9.10.1	**Introduction**	281
9.10.2	**Barren Worlds as Null Hypotheses**	283
9.10.2.1	Moon	283
9.10.2.2	Venus	284
9.10.2.3	Mars	284
9.10.2.4	Asteroids	284
9.10.3	**Biosignatures of Earth**	284
9.10.3.1	Inflammable Atmosphere	285
9.10.3.2	Permanent Liquid Water	285
9.10.3.3	Carbon-Based Life	285
9.10.3.4	Spheroidal Weathering	288
9.10.3.5	Granitic Rocks	289
9.10.3.6	Convex Slope	289
9.10.3.7	Calcic Horizon	290
9.10.3.8	Argillic Horizon	290
9.10.3.9	Mollic Epipedon	291
9.10.3.10	Civilization	291
9.10.4	**Origin of Life on Earth**	291
9.10.4.1	Experimental Studies	292
9.10.4.2	Geologic Records	292
9.10.4.3	Soil	293
9.10.4.4	Sea	293
9.10.4.5	Deep-Sea Vent	293
9.10.4.6	Other Worlds	293
9.10.5	**Coevolutionary Histories**	294
9.10.5.1	Life and Air	294
9.10.5.2	Life and Water	295
9.10.5.3	Life and Soil	296
9.10.5.4	Life and Rocks	298
9.10.6	**Conclusions**	299
Acknowledgments		300
References		300

> When I hear the sigh and rustle of my young woodlands, planted with my own hands, then I know that I have some slight share in controlling the climate (Anton Chekhov, 1899).

9.10.1 Introduction

Earth is one of the oldest words in our language (from Old English 'eorðe') meaning both our planet (customarily capitalized in this sense) and the soil beneath our feet. Here, both ends of meaning are explored by emphasizing how small-scale biological and soil-forming processes have played a role in the very significant differences between our planet and other planetary bodies in our solar system. Such processes as the growth of trees are commonly regarded as mundane, in the sense of commonplace or trivial, but they can be mundane also in the more important sense of global, as apparently understood by Chekhov (1899). The media that elevate biological processes to global scope are air, water, and soil. Air and water are important to the coevolution of life and earth (Nisbet and Sleep, 2001), but the path less traveled will be followed here, presenting a scenario on coevolution of life and earth from a research perspective of ancient and modern soils.

Soil is Anglo-French word, introduced from the ancient Roman 'solum' (ground). Soil is so central to human existence that it is as difficult to define as 'home' and 'love.' To some soil scientists, soil is material altered by the action of organisms, especially plant roots. To engineers, soil is material moved by mechanized graders without recourse to blasting. To farmers, soil is the surficial organic layer that nurtures crops and feeds livestock (Hole, 1981). A wider definition of soil is adopted here: materials on the surface of a planetary body altered in

place by physical, biological, or chemical processes. This is also the definition of NASA (National Aeronautic and Space Administration of the United States) press releases on lunar and Martian soils, for the compelling reason that they wished to be understood by the general public. In any case, the surface of the Moon and Mars is now altered locally by robotic landers and footprints, and future alterations can be anticipated. A broad definition of soil obviates the need for such terms as regolith or the distinction between physicochemical weathering and biological weathering (Taylor and Eggleton, 2001). More important to our present purpose, there is no longer an assumption that life created soil. The questions now become when and how did life evolve in soil? Did soil nurture the origin of life, or did life evolve in the ocean or volcanic hot springs and then later adapt to soil? These questions remain open (Nisbet and Sleep, 2001).

When soil, air, and waters became alive, unlike the apparently sterile but discernibly altered soils of the Moon, Venus, and Mars (**Figure 1**), it must have been an important time in the history of our Earth. Soils gave life access to fundamental nutrients such as phosphorus at their source in the weathering of rocks. Soils allowed life to colonize a significant fraction of Earth's surface, thus enhancing their volumetric effect on surficial fluids. Life gave soils increased depth and stability to enhance rates of physicochemical weathering. Life also gave soils a variety of biosignature horizons and deep weathering products (Retallack, 2001a).

The evolution of life and soil can be viewed as a coevolutionary process, like the coevolution of grasses and grazers, as first proposed by Kowalevsky (1873). Coevolution is the coordinated evolution of phylogenetically unrelated organisms, in this case plants and animals, which coevolved to enhance their mutual interdependence. Grasses evolved subterranean rhizomes, basal tillering of adventitious roots, intercalary meristems, telescoped internodes, and opal phytoliths to withstand grazing (**Figure 2**). Horses evolved cursorial limb structure for predator escape over open county and high-crowned teeth to withstand abrasion from leaves studded with opal phytoliths (MacFadden, 1992). Grasses and grazers coevolved to produce ecosystems, such as grasslands, and soils, such as Mollisols, that became biotic planetary forcings of the atmosphere and hydrosphere (Retallack, 2001b). Because coevolution is directed more toward other organisms than to physicochemical environments, Earth's environment coevolved with life far from primordial conditions exemplified by other planetary bodies (**Figure 1**).

The coevolution of life and soil is evident from functional intricacies of modern ecosystems, but the details of when and how this happened are shrouded in the mists of time. It is still uncertain, indeed largely hypothetical, when and how life first came to soils. There were changes in atmospheric oxidation

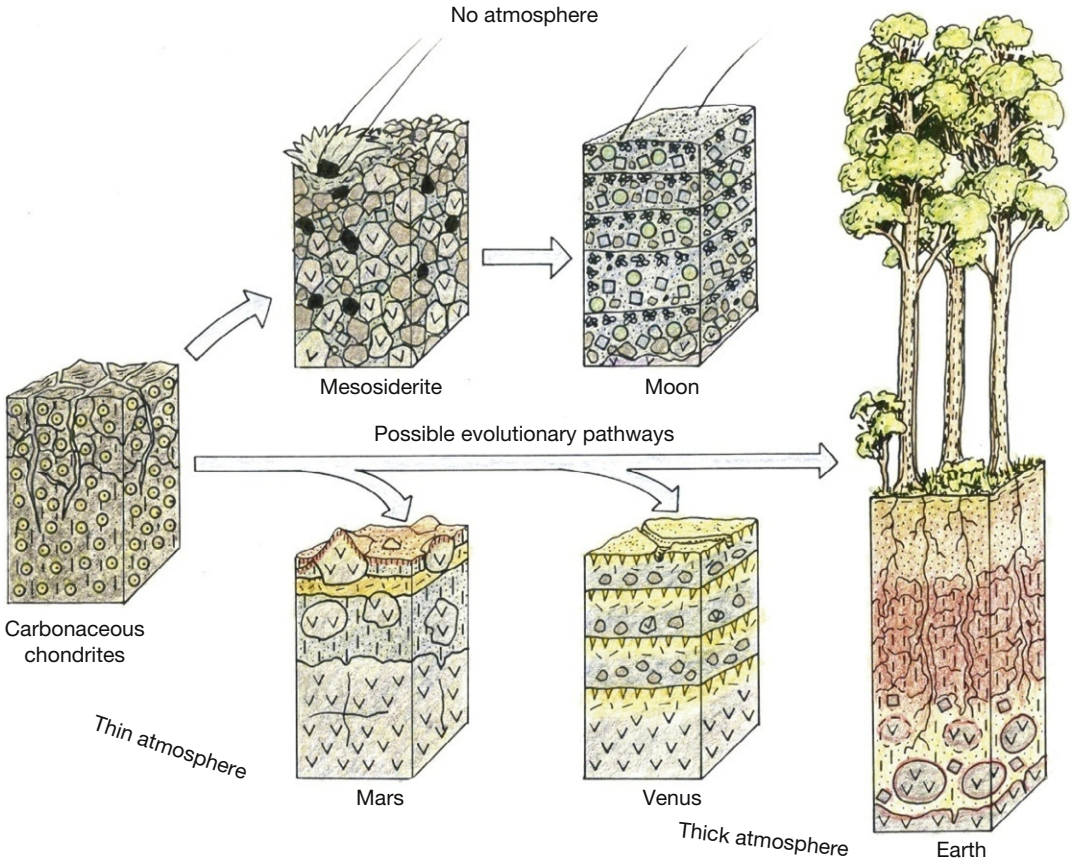


Figure 1 A schematic comparison of soil formation on Earth, Moon, Mars, Venus, and some hypothetical meteorite parent bodies. Reproduced from Retallack GJ (2001) *Soils of the Past*, 2nd edn., p. 600. Oxford: Blackwell, with permission.

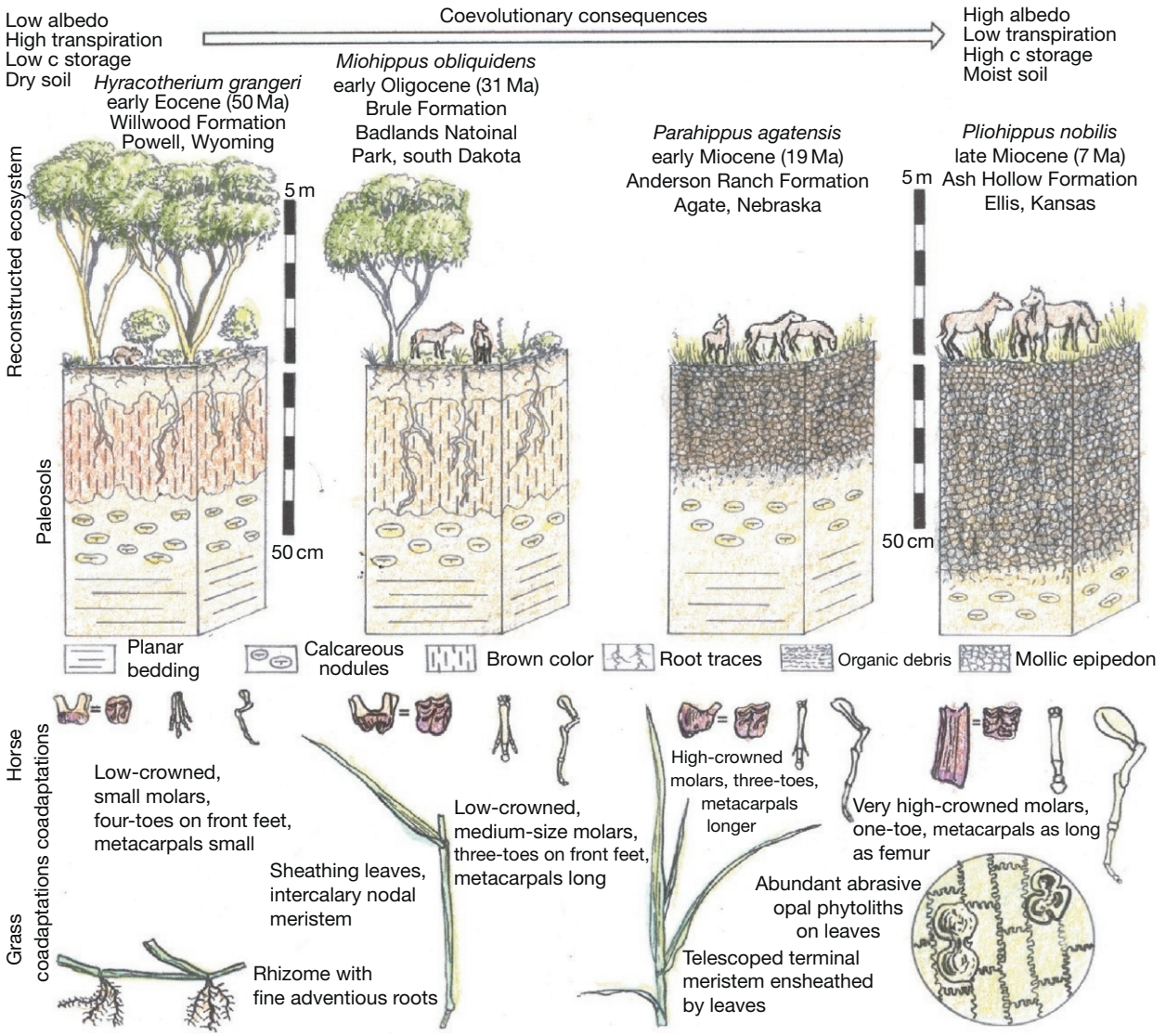


Figure 2 An example of coevolution, grasses and grazers, exemplified by paleosols and fossil horses from North America: grass evolutionary stages are less well constrained by the fossil record (data from Retallack, 2001b).

and water salinity on the early Earth, but it remains uncertain which of these are due to soil colonization, as opposed to oceanographic, volcanic, or tectonic changes. Fortunately, there is a remarkable record of paleosols in very ancient rocks on Earth (Rye and Holland, 1998) and on different planetary bodies (Retallack, 2001a), which furnish evidence on past conditions of climate and life. This supplements the geologic record of life and environments in marine, volcanic, and metamorphic rocks (Schopf, 1983; Schopf and Klein, 1991). Paleosol development and geochemistry also provide relevant tests for competing hypotheses of the development of continents and plate tectonics.

9.10.2 Barren Worlds as Null Hypotheses

Uniqueness of our planet is best considered by comparison with other planetary bodies, which turn out to have a variety of common features, such as basalt and CO_2-rich atmosphere. Water beyond Earth is frozen or vaporized, with the possible exception of Jupiter's moon Europa (Marion et al., 2005). Impact glasses, shocked minerals, clays, evaporites, carbonates, oxides, and a variety of high-temperature minerals also are found beyond Earth (Squyres et al., 2004). Also found beyond Earth are rugged desertlike landscapes, volcanoes, and ice caps (Carr, 1981). Such pervasive similarities to Earth make the search for biosignatures and their coevolutionary effects difficult, but before turning to that topic, consider some other worlds currently considered barren of life as null hypotheses.

9.10.2.1 Moon

At a small fraction of Earth mass (0.01), our Moon has insufficient gravitational pull to retain an atmosphere. Its temperature fluctuations of 111 to −171 °C also rule out liquid water (Taylor, 1982). Its soil appears waterless (Campbell et al.,

2006), despite past proposals of small areas of polar ice (Feldman et al., 1998). The main soil-forming process on the moon is micrometeoroid bombardment, which enriches basaltic breccias and rocks exposed at the surface in meteoritic iron–nickel, impact glass, agglutinates of glass and minerals, and shattered mineral grains (**Figure 1**). Well-developed lunar soils 5–10 cm thick take some 200 million years to form, judging from rate of micrometeoroid bombardment (McKay and Basu, 1983) and the succession of paleosols in the radiometrically dated Apollo 15 core (Heiken et al., 1976). The most profound soil-forming event in the history of the Moon was the footprints, tracks, and excavations of Apollo astronauts.

9.10.2.2 Venus

Venus is slightly smaller (0.8) than the mass of Earth, but its thick (93 bars) atmosphere is a toxic mix of greenhouse gases (97% CO_2) creating surface temperatures of about 470 °C. Water is not stable in the soil and water vapor is rare (0.1–0.4%) in the atmosphere. Soil formation on the basaltic parent material may be a rapid process akin to pottery glazing (**Figure 1**), in which salts are melted to obscure the hard edges of rocks and sediments (Barsukov et al., 1982).

9.10.2.3 Mars

Venus and our Moon stretch concepts of soil formation almost to breaking point, but Mars has more familiar ground. Mars is small (0.1 Earth mass), with a thin (0.008 bars) atmosphere, mainly CO_2 (95%), but with small amounts of N_2 (3%) and Ar (2%). Trace amounts (up to 0.003%) of CH_4 show temporal and geographic fluctuation suggestive of higher local concentrations (Formisano et al., 2004). The Viking lander at Chryse Planitia recorded temperatures from −90 °C at dawn to −30 °C at noon (Carr, 1981). Martian soils are thin with smectite clays and sulfate minerals (gypsum and kieserite) that indicate that basalt and komatiite parent materials have been altered in aqueous solution by hydrolysis (incongruent dissolution in carbonic acid). This kind of soil formation is slight compared with millions of years for similar soil formation in the Dry Valleys of Antarctica (Campbell and Claridge, 1987) or the Atacama Desert of Chile (Navarro-González et al., 2003). Soil formation may be even slower on Mars, because known soils appear to be paleosols from at least 3.2 Ga, formed at a time of more plentiful water at the surface indicated by outflow channels and then subsequently frozen at the surface during cold and dry subsequent eons (Retallack, 2001a). The red color of Martian soils is a very thin (<10 μm) rind, probably created by long-term surficial photooxidation (Yen et al., 2000). Mars has a stunning array of familiar landscapes including large volcanoes, deep canyons, impact craters surrounded by solifluction flows, eolian dunes, and evaporite basins (Squyres et al., 2004). Lifelike microstructures, organic chemicals, magnetites, and carbonates from a Martian meteorite found in Antarctica (McKay et al., 1996) raised hopes as microfossils, implying life on Mars as recently as 3.9 Ga. Although these remains are suggestive, they are not compelling, and Mars is widely considered sterile, now and in the geologic past (Schopf, 1999).

9.10.2.4 Asteroids

The soils of asteroids have only been visited recently (Veverka et al., 2001) and are brecciated by impact like lunar soils and some rare kinds of meteorites (mesosiderites and howardites; Bunch, 1975). Very different asteroidal soils are represented by carbonaceous chondrite meteorites, which show no evidence of living creatures, but include complex carbonaceous molecules, such as amino acids, dicarboxylic acids, sulfonic acids, and aliphatic and aromatic hydrocarbons (Pizzarello et al., 2001). These carbonaceous compounds are bound up within fine-grained phases of smectite and other clay minerals formed at low temperatures (Tomeoka and Busek, 1988) in contrast to high-temperature pyroxene and olivine and other high-temperature minerals of the chondrules (Itoh and Yurimoto, 2003). The clays form gradational contacts like weathering rinds, pseudomorphic replacement of chondrules, concentric cavity linings, and birefringence fabrics of oriented clay like those formed in soils under low confining pressures (Bunch and Chang, 1980). Other evidences of hydrous alteration are framboids and plaquettes of magnetite (Kerridge et al., 1979). There also are crosscutting veins of calcite, siderite, gypsum, epsomite, and halite (Richardson, 1978).

Clay and salts are common products of hydrolytic weathering in soils of Earth and can thus be inferred for the parent bodies of carbonaceous chondrites, which are thought to have been small (<500 km diameter) planetesimals. High-temperature chondrules of carbonaceous chondrites are 4.56 Ga in age (Amelin et al., 2002), and the plaquettes of magnetite (Lewis and Anders, 1975) and veins of calcite (Endress et al., 1996) and halite (Whitby et al., 2000) no younger than 4.51 Ga. Degassing of CO_2 and H_2O vapor from planetesimals early in the accretion of the solar system resulted in surprisingly Earthlike weathering (**Figure 2**).

This account has stressed a weathering interpretation of carbonaceous chondrites, but some of these also include talc and other minerals suggestive of high-temperature (250–300 °C) hydrothermal alteration (Brearley, 1997). There is also a school of thought that the clay and organic matter of carbonaceous chondrites are not alteration products, but instead cold condensates from the original nebula from which the solar system formed (Wasson, 1985).

9.10.3 Biosignatures of Earth

Several unique features of Earth compared with other planetary bodies can be considered biosignatures: features indicative of life. Some of these such as inflammable atmosphere, permanent liquid water, carbon-based life, and civilizations feature prominently in the search for life elsewhere in the universe. Biosignatures of Earth's soils include calcic, argillic, and mollic horizons. Salic horizons are known from Mars and carbonaceous chondrites. Carbon-rich surface horizons may be represented by carbonaceous chondrites. It is not yet clear whether deep hydrolytic weathering occurs on other planetary bodies (Squyres et al., 2004). Other features as convex slopes, spheroidal deep weathering, and granitic rocks unique to Earth are less obviously related to life's influence, but nevertheless worth consideration.

9.10.3.1 Inflammable Atmosphere

Our atmosphere is a combustible mix of oxygen (21%) and methane (0.00017%), with its reacted product carbon dioxide (preindustrial 0.028%, but 0.0397% in 2012 and rising), diluted with much inert nitrogen (78%). It is thus more like input gases of a carburetor than the exhaust gases of an automobile tailpipe and quite distinct from the CO_2-rich composition of other planetary atmospheres and volcanic gases (Lovelock, 2000). Burning of other flammable substances such as hydrocarbons and wood in the open air is thus possible on our planet, unlike other planetary bodies. More than trace amounts of oxygen in the atmosphere date back at least 2.3 Ga (Bekker et al., 2004). Entropy reduction or chemical disequilibrium of planetary atmospheric composition is a way of identifying life elsewhere in the universe and also an argument for the Gaia hypothesis of life's control of atmospheric composition (Lovelock, 2000).

The reason for the cosmically peculiar composition of our atmosphere may be the widespread metabolic process of photosynthesis (Rosing et al., 2006): not only by plants, and a variety of protists, but also by a host of bacteria. Photosynthesis, or assembly by light, is a process by which sugars are synthesized using light from the sun and a catalyst (commonly chlorophyll) from CO_2 of the atmosphere (Table 1). The low amount of CO_2 in our current atmosphere is a testament to the success of photosynthesis, but replacement of a primordial reducing atmosphere of CO_2 and CH_4 by O_2 had many ups and downs over 3.5 Ga of Earth history known from the rock record (Berner et al., 2000).

9.10.3.2 Permanent Liquid Water

Our pale dot of a planet, blue with oceans and wreathed with water vapor, is unique in our solar system now and also in its deep history of water at the surface (Lovelock, 2000). On a cosmic scale of temperatures from −270 °C (2.76°K) in deep space to about 6000 °C for the surface of the sun, the 0–100 °C range of liquid water is a narrow range not currently found on any neighboring planetary bodies, though perhaps present within Europa (Marion et al., 2005), and on Mars in the distant geologic past, some 2.5–3.5 Ga (Carr, 1981). It did not last on Mars, but on Earth, there is a continuous record of water-lain sedimentary rocks and precipitates from 3.5 Ga to the present (Allwood et al., 2006). Oxygen isotopic composition of detrital zircons individually dated to 4.3 Ga in the Jack Hills of Western Australia is unusually heavy (15.3 vs. typical mantle values of 5.4‰ $\delta^{18}O$ vs. SMOW) and suggests free surficial water at the time their parent granitoids cooled only 200 Ma after the formation of Earth (Mojzsis et al., 2001). This is a long and continuous history of surficial water despite degassing of largely greenhouse gases to a thick (1 bar) atmosphere and a 30% increase in solar luminosity due to stellar evolution over the same period. Permanent water and its longevity through planetary history are a clue to life on other planets, as well as an argument for Gaian environmental regulation (Lovelock, 2000).

The reason for the maintenance of Earth's temperature within the bounds of liquid water for some 4.3 Ga remains uncertain, but life is suspected to have a role because of its coevolved metabolic systems. This general idea of a thermocouple of opposing forces acting as a thermostat is clearly expressed in 'Daisyworld' albedo models of the Gaia hypothesis (Watson and Lovelock, 1983). Imagine a world with only black and white daisies vying for the light of the sun in order to photosynthesize. White daisies cool by reflecting sunlight back into space. Black daisies warm by absorbing heat. Populations will be dominated alternately by black and then white daisies, until mixed populations converge on temperatures that optimize photosynthesis. A more realistic model is the 'Proserpina principle' (Retallack, 2004a), which postulates that photosynthesis cools the planet by drawing down the greenhouse gas CO_2, whereas respiration warms the planet because it draws down O_2. If either one of these metabolic systems had evolved in isolation, it would have resulted in a respirator's hell like Venus or a photosynthesizer's freezer like Mars. Fortunately, photosynthesizers and respirers, such as plants and animals, are mutually interdependent for food and breath: a coevolutionary thermostat. Plants cool the planet by photosynthesis, which is curtailed as they are covered by snow of icehouse atmospheres. Animals warm the planet by respiration but die in high temperatures of greenhouse atmospheres that are less fatal to plants. Population balances between them have the effect of adjusting greenhouse gases in the atmosphere to habitable ranges given external inputs of solar radiation.

9.10.3.3 Carbon-Based Life

Science fiction movies make alien life instantly recognizable by emphasizing the salient features of humans or insects. Recognizing alien life on sight may not be so easy because much life is immobile or microscopic, and its activities too varied to be easily characterized (Table 2). Life on Earth has a marked

Table 1 Metabolic processes of geologic significance

I. Fermentation	$\underset{\text{sugar}}{C_6H_{12}O_6} \rightarrow \underset{\text{alcohol}}{2C_2H_5OH} + \underset{\text{carbon dioxide}}{CO_2} + \text{energy}$
II. Respiration	$\underset{\text{sugar}}{C_6H_{12}O_6} + \underset{\text{oxygen}}{O_2} \rightarrow \underset{\text{carbon dioxide}}{6CO_2} + \underset{\text{water}}{6H_2O} + \text{energy}$
III. Anaerobic photosynthesis	$\underset{\text{carbon dioxide}}{6CO_2} + \underset{\text{hydrogen sulfide}}{12H_2S} \xrightarrow[\text{bacteriophyll}]{\text{light}} \underset{\text{sugar}}{C_6H_{12}O_6} + \underset{\text{water}}{H_2O} + \underset{\text{sulfur}}{12S}$
IV. Aerobic photosynthesis	$\underset{\text{carbon dioxide}}{6CO_2} + \underset{\text{water}}{12H_2O} \xrightarrow[\text{chlorophyll}]{\text{light}} \underset{\text{sugar}}{C_6H_{12}O_6} + \underset{\text{water}}{6H_2O} + \underset{\text{oxygen}}{6O_2}$

Table 2 Kinds of microbes, their metabolic requirements and role

General role	*Kind of organism*	*Example genus*	*Energy source*	*Electron donor*	*Carbon source*	*Oxygen relations*	*Comments*
Carbon cyclers	Algae	*Chlamydomonas*	Sunlight (P)	H_2O (L)	CO_2 (A)	Aerobic	Primary producer normal–wet soils, ponds, ocean
	Cyanobacteria	*Nostoc*	Sunlight (P)	H_2O (L)	CO_2 (A)	Aerobic	Primary producer normal–wet soils
	Purple non-sulfur bacteria	*Rhodospirillum*	Sunlight (P) sometimes organic compounds (C)	Organic compounds (O) sometimes H_2S (L)	Organic compounds (H) sometimes CO_2 (A)	Amphiaerobic	Primary producer in poorly oxygenated soils, stagnant ponds, and ocean
	Methanogenic bacteria	*Methanobacterium*	H_2 (C)	H_2 (L)	CO_2 (A) sometimes formate (H)	Anaerobic	Creators of 'swamp gas' (CH_4) in stagnant ponds and ocean
	Aerobic spore-forming bacteria	*Bacillus*	Organic compounds (C)	Organic compounds (O)	Organic compounds (H)	Aerobic	Decomposer of organic compounds in soils, lakes, and oceans
	Fermenting bacteria	*Clostridium*	Organic compounds (C)	Organic compounds (O)	Organic compounds (H)	Anaerobic	Decomposer in poorly oxygenated soils, lakes, and oceans
	Protoctistans	*Amoeba*	Organic compounds (C)	Organic compounds (O)	Organic compounds (H)	Aerobic	Predator in soils, lakes, and oceans
Nitrogen cyclers	Nitrogen-fixing bacteria	*Azotobacter*	Organic compounds (C)	N_2 (L)	Organic compounds (H)	Aerobic	Creates ammonium (NH_4^+) in soils, lakes, and oceans
	Root nodule bacteroids	*Rhizobium*	Organic compounds (C)	N_2 (L)	Organic compounds (H)	Aerobic	Supplies ammonium (NH_4^+) to host plant in well-drained soils
	Denitrifying bacteria	*Pseudomonas*	NO_2 (C) sometimes organic compounds (C)	N_2 (L) sometimes organic compounds (O)	CO_2 (A) sometimes organic compounds (O)	Anaerobic	Releases nitrogen (N_2) to atmosphere from poorly oxygenated soils
Sulfur cyclers	Sulfur-reducing bacteria	*Desulfovibrio*	Organic compounds (C)	SO_4^{2-} (L)	Organic compounds (H)	Anaerobic	Creates 'rotten egg gas' (H_2S) and pyrite in swamps, lakes, and oceans

(*Continued*)

Table 2 (Continued)

General role	*Kind of organism*	*Example genus*	*Energy source*	*Electron donor*	*Carbon source*	*Oxygen relations*	*Comments*
	Sulfur bacteria	*Chromatium*	Sunlight (P)	H_2S or S (L) sometimes organic compounds (O)	CO_2 (A) sometimes organic compounds (H)	Amphiaerobic	Remobilizes sulfur in poorly oxygenated soils, lakes, and oceans
	Sulfur metabolizing bacteria	*Thiobacillus*	S, FeS (C)	S, SO_4^{2-}, Fe_2O_3 (L)	CO_2 (A)	Aerobic	Remobilizes sulfur in normal–wet soils, lakes, and oceans

Initials for trophic groups are autotrophic (A), chemotrophic (C), heterotrophic (H), lithotrophic (L), organotrophic (O), and phototrophic (P).
Source: Pelczar MJ, Chan ECS, Krieg NR, and Pelczar ME (1986) *Microbiology*, p. 918. New York: McGraw-Hill.

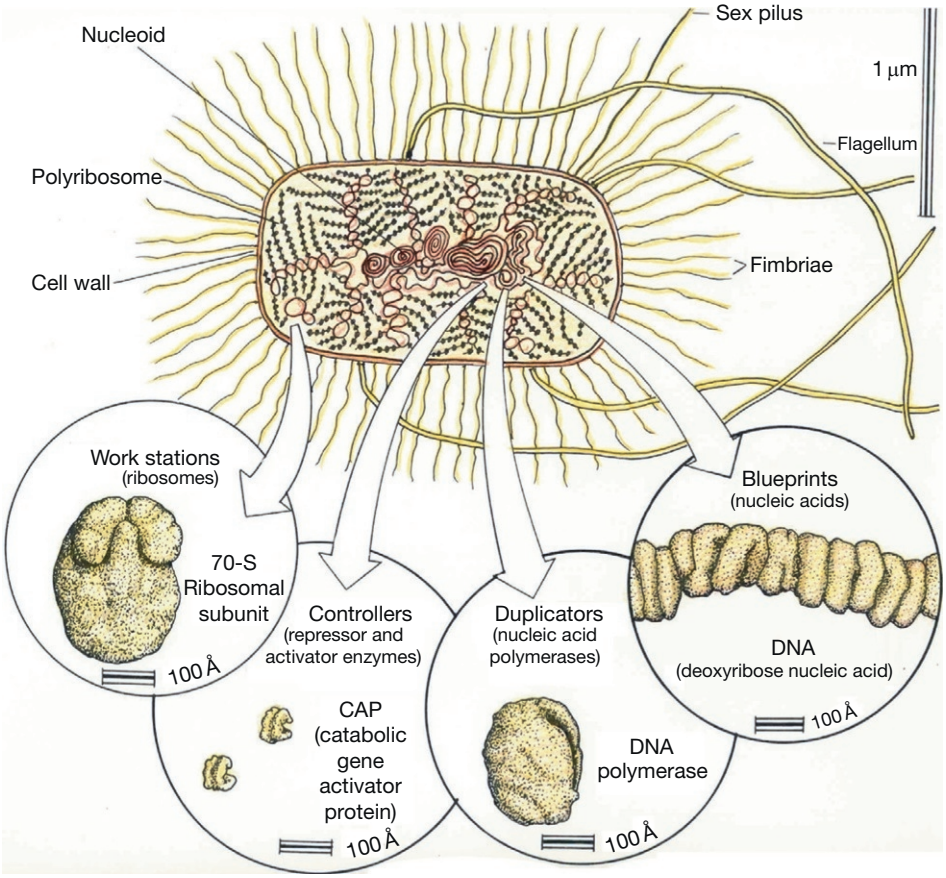


Figure 3 *Escherichia coli*, a common gut bacterium, considered as an organic machine of astonishing complexity. Reproduced from Retallack GJ (2001) *Soils of the Past*, 2nd edn., p. 600. Oxford: Blackwell, with permission.

preference for six elements: carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur (CHONPS; Schoonen et al., 2004). Of these, only oxygen is a common component of most rocks. Three general features are regarded as necessary for life: bodies, metabolism, and reproduction. Even bodies of such simple organisms as the common gut bacterium *Escherichia coli* have an astonishing complexity of interacting parts, including high-molecular-weight organic compounds that provide both structure and function (Figure 3). If there is a common theme to the great variety of metabolic reactions, it is lack of chemical equilibrium that drives them and that has left an imprint on our cosmically peculiar atmosphere (Lovelock, 2000). Reproduction is also quite varied, ranging from simple cell division, to clonal budding, to parthenogenesis, to the romantic complexity of sex. If there is a common theme to biological reproduction, it is remarkable speed. Astonishing

numbers of individuals can be produced in days or years as long as food is available. Whole worlds can be populated in a geologic instant, which is encouraging to the search for life on other worlds. If life is anywhere on a planet, it is likely to be everywhere.

The difficulties of recognizing life are well illustrated by debate over evidence of life on the early Earth and a Martian meteorite. Very ancient photosynthetic activity is suggested by the carbon isotopic ratios of amorphous organic matter in the 3.8 Ga Isua Greenstone Belt of west Greenland (Rosing and Frei, 2004), although the high metamorphic grade and discernable modern organic contamination are concerns (Nishizawa et al., 2005). The oldest microfossil evidence for life has long been considered chains of cyanobacterium-like carbonaceous cells in chert from the 3.5 Ga Apex Chert of the Warrawoona Group near Marble Bar, Western Australia (Schopf, 1983). The poor preservation and hydrothermal setting of these fossils have recently been urged as evidence that they are not microfossils, but inorganic carbonaceous material (Brasier et al., 2002). In my opinion, the distortion and poor preservation of the microfossils, and especially the rounded cavities in them (Schopf et al., 2002), are evidence of both cyanobacterial primary production and actinobacterial decay. Carbonaceous and mineral pseudofossils have greater regularity of construction and resistance to degradation than the Warrawoona microfossils (Garcia-Ruiz et al., 2003). In any case, alternative evidence of life from microbial etching of pillow laves of the Barberton greenstone belt of South Africa is also dated to about 3.5 Ga (Furnes et al., 2004). Furthermore, a variety of distinctive microbial textures are now known from the 3.58 Ga Dresser Formation of Western Australia (Van Kranendonk et al., 2008).

The Martian meteorite (ALH84001) with putative microfossils recovered from near Allan Hills, Antarctica (McKay et al., 1996), is of an unusual type (Shergotty-Nakhla-Chassigny three representative meteorites (SNC)) thought to have been derived from Mars because of its igneous (nonchondritic) microtexture, geologic age younger than the 4.6 Ga age of most meteorites, unusual oxygen isotopic composition, lack of magnetic paleointensity, and evidence for long exposure to interplanetary cosmic rays (Wood and Ashwal, 1981). Meteorite ALH84001 groundmass is 4.5 Ga in age but was shock metamorphosed at 4.0 Ga and then infiltrated by secondary carbonate with putative microfossils at 3.9 Ma (Borg et al., 1999). It was lofted from Mars by about 16 Ma and orbited the sun until falling into Antarctic ice at about 13 Ka (McKay et al., 1996). The putative microfossils are cylindrical shapes arranged end to end in short strings, with associated polycyclic aromatic hydrocarbons and lozenges of magnetite comparable with magnetotactic bacteria and nannobacteria on Earth (Folk and Lynch, 1997). Unfortunately, the cell-like structures are too small (0.02–0.1) to house the complex molecular machinery of even the simplest known bacteria (**Figure 3**) and are more likely organic coagulates or artifacts of scanning electron microscopy (Schieber and Arnott, 2003). The organic compounds have isotopic composition (Jull et al., 1998) and racemization like Earthly contaminants (Bada et al., 1996). The magnetite has nonbiological screw dislocations and the carbonate may have formed at high temperatures (McSween, 1997). Life on Mars thus remains unproven.

9.10.3.4 Spheroidal Weathering

Spheroidal weathering is the deep weathering of rocks of uniform composition, especially granitic rocks, to create rounded concentric rings of alteration inward from deep cracks (**Figure 4(a)**). Onion-skin weathering is a particularly aggressive form of spheroidal weathering in which the whole rock has been altered to concentric shells. More common is a few shells of alteration around a rounded remnant of little altered rock in the center, a corestone. Rounded corestones resist erosion more strongly than soft surrounding rock and overlying soil profile (Strudley et al., 2006) and are exhumed as distinctive rounded boulders called tors (**Figure 4(b)**). No landforms or deep weathering of these kinds has yet been found on the Moon, Venus, Mars, or asteroids, but this kind of deep weathering is common in deeply weathered Precambrian paleosols as old as 2.3 Ga (**Figure 4(a)**). If spheroidal weathering has such geologic antiquity, one would expect tors

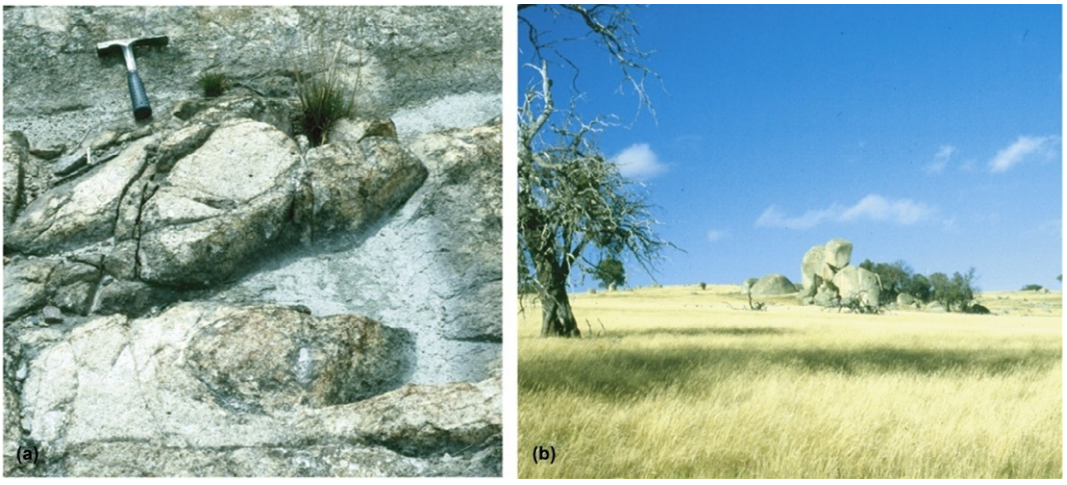


Figure 4 Spheroidal weathering and tors. (a) Paleoproterozoic (2.3 Ga) spheroidally weathered corestones in a paleosol developed on pink alkali granite and covered Huronian Supergroup conglomerates, near Elliot Lake, Ontario, Canada (hammer gives scale). (b) Tors of Baynton Granite exhumed from Miocene paleosol, 4 km north of Pyalong, Victoria, Australia (tree is 5 m tall).

to be ancient landforms as well, but the oldest likely documented case known to me is on the pre-Stoer (1.2 Ga) paleosol near Stoer, northwest Scotland (Stewart, 2002).

Spheroidal weathering and exhumed tors require deep weathering by acidic solutions that are promoted by abundance of water and warmth (Taylor and Eggleton, 2001). Strongly acidic solutions of the sort that form during oxidation of pyritic mine spoils are not indicated, because these form intense localized plumes of alteration. More likely, the acid involved in spheroidal weathering is carbonic acid formed from dissolution of CO_2 in soil water exploiting the network of cracks in the rock due to expansion by thermal stresses or following unloading of overburden (Retallack, 2001a). Warm, wet carbonic solutions should be feasible on other planetary bodies, but on Earth, they are greatly promoted by life. Soil respiration, especially of microbes deep in soil profiles, creates partial pressures of CO_2 up to 110 times that of atmospheric pressures of CO_2 on Earth (Brook et al., 1983). Texturally uniform rocks, such as anorthosites on the Moon and komatiites on Mars, are produced on other planetary bodies, though not granitic rocks (Campbell and Taylor, 1983).

9.10.3.5 Granitic Rocks

Granitic rocks are unique to Earth and not yet discovered on other planetary bodies or meteorites (Campbell and Taylor, 1983). Granitic rocks are ancient but not primordial, with granodiorites appearing at about 3.0 Ga among more abundant tonalites and trondhjemites of continental crust (Engel et al., 1974). They thus postdate evidence for life (Schopf, 1983), water (Mojzsis et al., 2001), and deep hydrolytic weathering on Earth (Buick et al., 1995), so it is plausible that granites are biosignature rocks.

It may seem outrageous to suggest that granitic rocks forming under high temperatures (600–800 °C) at depths of 2–10 km within continental crust could be in any way related to life, but additional features beyond uniqueness to Earth and geologic antiquity suggest a role for life in granitization (Rosing et al., 2006). Granites are low density, hydrated, aluminous, and siliceous compared with other igneous rocks, and these are all qualities imparted by deep hydrolytic weathering known to have been widespread on Earth back some 3.5 Ga (Maynard, 1992). Light rare earth enrichment, including preference for neodymium over samarium, is also a geochemical similarity of granitization and weathering (MacFarlane et al., 1994). The 'Daly Gap,' or rarity of rocks intermediate in composition between basalt and rhyodacite, in many continental volcanic regions points to different pathways of genesis (Bonnefoi et al., 1995). Migmatites and S-type granites include melted sedimentary rocks derived from continental weathering, judging from their elevated $^{86}Sr/^{87}Sr$ ratios, oxygen isotope composition, and alumina content, as well as minerals such as corundum, muscovite, garnet, and cordierite (Rosing et al., 2006). By this view, life's role in forming granite begins with its promotion of deep weathering and surficial liquid water. Some of this low-density, aluminous, and siliceous soil and sediment would be melted upon deep burial by overlying rocks or entrained by the foundering of dense basaltic slabs in plate tectonic recycling to form a new kind of rock, granite. From this perspective, it is also plausible that plate tectonics driven by density differences and surficial water leading to granitization is also a biosignature of Earth. There is no plate tectonics on the Moon (Taylor, 1982), and it seems unlikely on Venus, which has some poorly understood mechanism of crustal renewal (Fowler and O'Brien, 2003), but early Mars may have had seafloor spreading (Nimmo and Tanaka, 2005).

A role for life and weathering in granitization is rarely considered in igneous petrology. The prevailing view of granite formation is by magmatic processes, such as fractional crystallization in which mafic minerals drop out of molten magmas to form dense residues leaving more silicic fluids to form granitic rocks (Rudnick, 1995). Remarkably, Bowen's (1922) famous reaction series, which defines the order in which specific minerals crystallize from silicate melt, is the same order in which they are destroyed in deeply weathered soils, as determined by Goldich (1938). Olivine and pyroxene are early to crystallize in magmas and first to be weathered; in contrast, quartz is last to crystallize and weather away.

9.10.3.6 Convex Slope

Concave and straight slopes are common in desert badlands and alpine glacial terrains on Earth, the Moon, and Mars. Broadly convex slopes on the other hand may be unique to soil-mantled upland topography on Earth (Dietrich and Taylor Perron, 2006). The effect is subtle in topographic profiles, especially those with low topographic relief, but can be striking in slope-area plots of small areas (<6 km^2) compiled from high-resolution digital elevation models (**Figure 5**). Soil-mantled terrain has low slopes around upland small drainage basins but steep slopes in large valleys, whereas badlands and mountains show comparably steep slopes in both small and large drainage basins. Put another way, landscapes at large scale and with limited life are almost fractal in distribution of slope with catchment size, but soil-mantled landscapes are nonfractal in small catchments. This potential biosignature deserves further attention through image analysis of other planetary bodies and paleosols.

Concave slopes are created by fluvial or other fluid erosion, by Gilbert's (1877) 'law of divides' in which slope increases with restricted discharge upstream. Straight slopes on the other

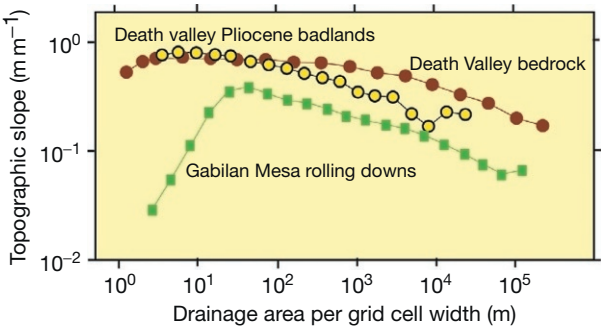


Figure 5 Slope–area relations in soil-mantled compared with desert bedrock and badlands in landscape areas of 0.24–5.8 km^2 in California (reproduced from Dietrich WE and Taylor Perron J (2006) The search for a topographic signature of life. *Nature* 439: 411–418, with permission). Drainage area per grid cell is the size of individual catchments within successively larger map squares beginning with 2 m squares.

hand are the angle of repose of scree and glide planes of other forms of mass movement, as first noted by Penck (1953). Broadly convex slopes in contrast are associated with thick soil mantles. Soil communities bind the earth on which they depend while smoothing with bioturbation irregularities in the overall expansion of parent materials due to weathering and cracking. Mathematically, this is diffusive dilation. Soil stabilized by life creeps downslope until failure angles are reached to create broad convex slopes above oversteepened straight to concave slopes. In contrast, desert badlands and alpine glacial terrains have narrowly convex divides, steep headwall drainages, and long straight to concave slopes.

9.10.3.7 Calcic Horizon

A calcic horizon is a particular level within a soil studded with nodules of calcite or dolomite (**Figure 6(a)**) and is especially common in desert soils (Aridisols) on Earth. Despite the desertlike images of the Moon (Taylor, 1982), Venus (Barsukov et al., 1982), and Mars (Squyres et al., 2004), no pedogenic carbonate has yet been detected beyond Earth. Spectroscopic detection of carbonate minerals in Martian dust (Bandfield et al., 2003) may have the same source as carbonate in Martian meteorites and carbonaceous chondrites, which are vein-filling crystalline carbonate, some of it formed at high temperatures (McSween, 1997). Fossil and modern pedogenic carbonate in contrast is micritic and replacive in microtexture (Retallack, 1991).

Many aspects of calcic horizons are controlled by their soil ecosystems. Calcareous rhizoconcretions and termitaries are biogenic structures modeled on roots and termite nests, respectively (Retallack, 2001a). Micritic replacement within nodules is microbially mediated (Monger et al., 1991). The depth to calcic horizons in soils is strongly correlated with mean annual precipitation and soil respiration (Retallack, 2005). Calcic horizons are found largely within the mean annual precipitation range of 300–1000 mm, as shown particularly well in transects from the hyperarid Atacama Desert to higher elevation shrubland and grassland (Rech et al., 2006). The microbe-poor Atacama Desert soils have gypsic horizons of soluble salts, like soils of Mars (**Figure 1**) and the Dry Valleys of Antarctica (Campbell and Claridge, 1987). Calcic horizons are found in soils of desert shrublands largely leached of such soluble salts, which inhibit biological activity with high osmotic stress. Another difference is the dominance of lichens in calcic, rather than mainly bacteria in gypsic soils, because magnesium and calcium oxalate secondary metabolites of lichens promote formation of pedogenic carbonate (Verrechia et al., 1993).

9.10.3.8 Argillic Horizon

An argillic horizon is a level within a soil markedly enriched in clay compared with horizons above and below (**Figure 6(b)**) and is a hallmark of forested soils such as Alfisols and Ultisols. The argillic horizon not only is defined on the basis of amounts of clay enrichment (roughly 10% more, with exact definition texture-dependent, Soil Survey Staff, 2010) but also has characteristic microstructure of extremely fine-grained and high-birefringence wisps of pedogenic clay along with

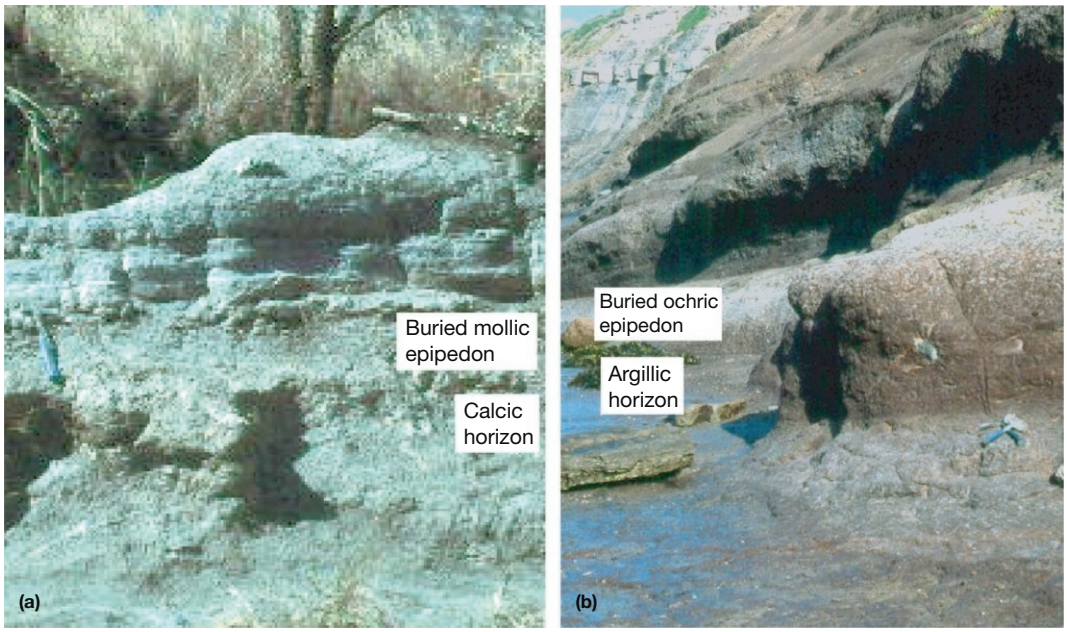


Figure 6 Calcic and argillic horizons and mollic epipedon in paleosols. (a) Two Chogo clay paleosols (Calciustolls) showing mollic epipedon (dark surface) and calcic horizon (white nodular bands), both profiles truncated by nephelinitic granular ash in the Middle Miocene Fort Ternan Member of the Kericho Phonolites, near Fort Ternan, Kenya. Reproduced from Retallack GJ (1991) *Miocene Paleosols and Ape Habitats of Pakistan and Kenya*, p. 246. New York: Oxford University Press. (b) Long Reef clay paleosol (Paleudult) showing argillic horizon (dark red clay-enriched) beneath light-colored surface horizon covered by clayey alluvium in the late Early Triassic, Bald Hill Claystone, near Long Reef, New South Wales, Australia. Adapted from Retallack GJ (1997) Palaeosols in the Upper Narrabeen Group of New South Wales as evidence of Early Triassic palaeoenvironments without exact modern analogues. *Australian Journal of Earth Sciences* 44: 185–201. Hammers give scale in both frames.

inherited clay (Retallack, 2001a). These kinds of horizons have not been found on the Moon, Mars, and Venus and are not found in paleosols on the early Earth either. The geologically oldest known example of an argillic horizon is from Middle Devonian (Givetian, 390 Ma) Aztec Sandstone of Victoria Land, Antarctica (Retallack, 1997a). Argillic paleosols and other evidences of ancient forests are common in every subsequent geologic period (Retallack, 2001a).

Both the micromorphology and geologic history of argillic horizons suggest that they are products of forest ecosystems, known independently from fossil stumps and wood to have originated by Middle Devonian (390 Ma). Their micromorphology of wispy to laminated very fine-grained clay suggests that they form from both washing down and weathering in place around large tapering roots (Retallack, 2001a). During interplanetary exploration, it is likely that forests would be encountered before argillic horizons, but relict argillic paleosols could be indicative of forests of the geologic past, just as they are in some desert regions of Earth today.

9.10.3.9 Mollic Epipedon

The mollic epipedon is a unique surface horizon of grassland soils (Mollisols), which presents a number of paradoxes. It is rich in base-rich clay such as smectite yet does not ball up or shear into unwieldy clods and flakes. It is rich in organic matter yet well aerated with cracks and channels in which organic matter should be oxidized. It can be well drained yet still maintains soil moisture. It is fertile with phosphorus and other mineral nutrients despite long periods (thousands of years) of soil development. Mollic epipedons achieve these highly desirable agricultural qualities through a unique structure of fine (2–3 mm) ellipsoidal (crumb) clods (peds), which consist of clay stabilized by organic matter. The geologically oldest fossilized mollic epipedons are Early Miocene (19 Ma) and known from the Anderson Ranch Formation, near Agate, Nebraska, and the upper John Day Formation, near Kimberly, Oregon, the United States (Retallack, 2004b). Nothing like a mollic epipedon has been reported from the Moon, Venus, Mars, or meteorites.

The temporal range and micromorphology of mollic epipedons suggest that they are created by grassland ecosystems. Some of the rounded crumb peds are excrements of earthworms common in these soils; other peds are products of the three-dimensional network of the slender adventitious roots of sod-forming grasses (Retallack, 2001a). The organic matter is partly from root exudates and earthworm slimes, but cake-like dung of large ungulates also plays a role in soil conditioning. The mollic epipedon can be considered a trace fossil of sod-grassland ecosystems, even after destruction by desertification or burial (Retallack, 2004b).

9.10.3.10 Civilization

One would think the Great Wall of China a human construction obvious from space. In poor resolution images, it could be confused with another 'Great Wall of China,' which is a tourist attraction south of Blinman in the northern Flinders Ranges, Australia, and a natural outcrop of Neoproterozoic Etina Formation flanking a diapiric uplift like a fortified city. The so-called face of Elvis and pyramids on Mars were subsequently shown to be inselbergs and yardangs (Pieri, 1999). Similarly, the canals of Mars were products of imagination working at the limits of optical resolution of the time (Zahnle, 2001). Close optical and physical surveillance will be needed to detect civilizations, even on planets with dead atmospheres and boiled oceans in which civilizations may once have failed. Successful civilizations on the other hand may come to us. The effort to detect radio communications from outer space continues with ever larger telescopes and more powerful computer power (Whitfield, 2003). Yet we remain alone and isolated in the universe and know no failed civilizations either.

9.10.4 Origin of Life on Earth

From a coevolutionary perspective, the origin of life is not the heroic struggle of an individual molecule or organism from a primordial soup (Nisbet and Sleep, 2001), but a system of supportive molecules or organisms for growth and replication in an environment that selected for some desirable property while metering the supply of fundamental nutrients to last for geologic timescales. To oversimplify in human terms, the first life needed a job, a family, a place to live, and a legacy for future generations. These theoretical requirements are made clear by Cairns-Smith (1971), who has argued that the precursors of organic life were clays that grew by assimilating cations from weathering solutions and reproduced by cracking into smaller pieces or flakes with diagnostic layering or cationic substitutions. These general characteristics would not coevolve in any lifelike way unless the clays also had some property subject to natural selection. He proposed the simple but important quality of stickiness. The layering and cationic substitution of smectite clays determine their interlayer expansion upon wetting, which can vary stickiness from a thin slurry to a thick paste. Stickiness becomes important because it can be selected to stabilize an environmental interface and thus create a legacy for creation of more stickiness to maintain that interface in a critical zone of energy and material flux.

Cairns-Smith thought in terms of a sandy aquifer, but his model works better in a primordial soil of a planetesimal, in which minerals are weathered by carbonic acid solutions of degassed CO_2 and rain and warmed by a faint young sun. Unlike Cairns-Smith's water-saturated model, this hypothetical primordial soil had millions of warm little ponds suspended between and within grains between rains (**Figure 7**). Sloppy clay expands violently when wet (like palygorskite) and is then washed too deep into the soil for warmth and acid to produce copies. Tough clay on the other hand expands little and covers source grains in a thick rind (like kaolinite) so that the mineral is no longer available to water and acid to produce more clay. Lumpy and sticky clay in contrast has a consistency that expands on wetting sufficiently to bridge the gaps between grains, yet crack away from the source grains to allow access for water and acid to create more clay (like smectites). These clays will be selected against the soil-eroding activities of wind and water because their grains are glued together by sticky clays. In an environment where loose sandy soils are eroded, all soils become more clayey and by persisting produce more clay. Lumpy clays may be better adapted to heavy rains, and

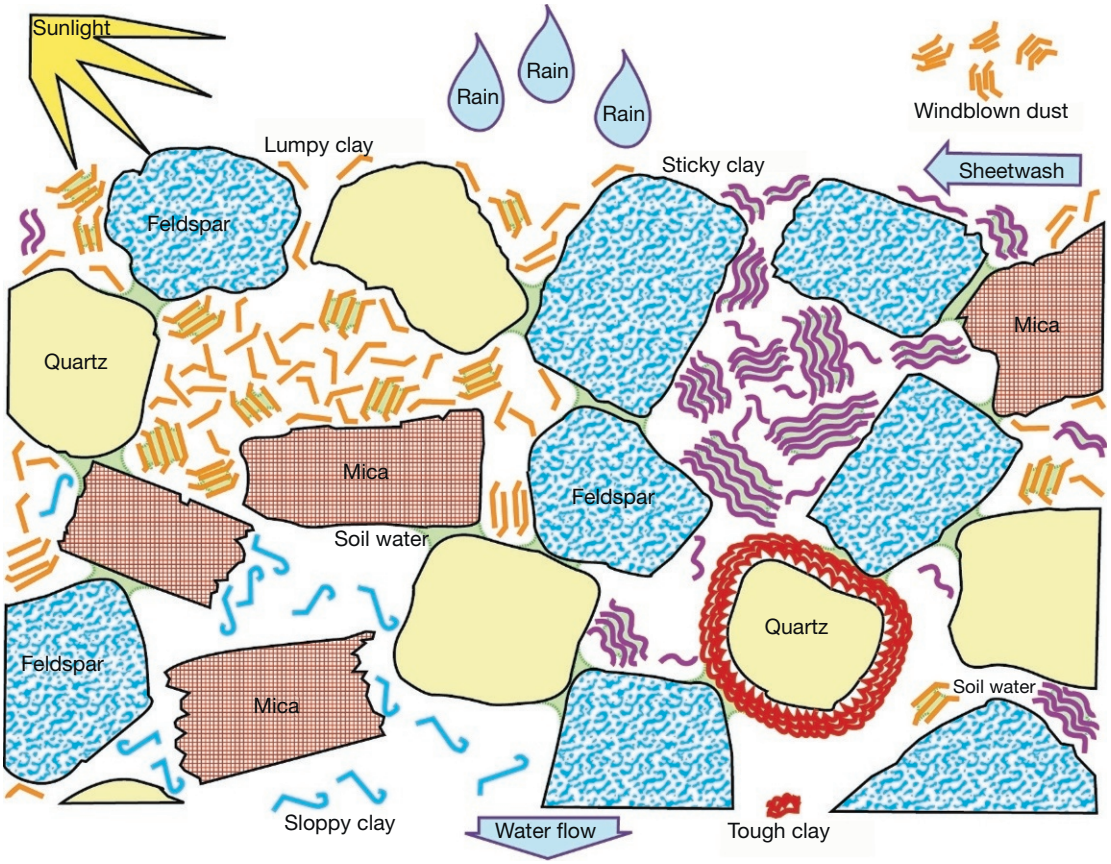


Figure 7 Hypothetical natural selection of four kinds of clay differing in stickiness (sloppy, lumpy, sticky, and tough) and weathering from the minerals of a primordial soil. Reproduced from Retallack GJ (2001) *Soils of the Past*, 2nd edn., p. 600. Oxford: Blackwell, with permission.

sticky clays to drizzle, but both are selected by their ability to hold the soil. They are in a coevolutionary escalation that refines their distribution, composition, and degree of stickiness for particular climatic conditions, because rivers erode the least sticky clays. Similarly, the first organic molecules, produced by abiotic reactions, could also be selected in soils for their stickiness. Like molasses on a cornfield, organic soils survive erosion to make more organic matter and create soils with observed fabrics of carbonaceous chondrites. The molecules of heredity, DNA and RNA, are sticky too and could have begun what Cairns-Smith calls a 'genetic takeover' of the older and cruder clay system of soil stabilization. Cells, trees, and civilizations continue to depend on the stability and bounty of soil and its nutrients. This simple parable makes clear the importance of natural selection by the environment in converting the long odds of chance evolution of complex life to a necessity.

9.10.4.1 Experimental Studies

Laboratory experiments have succeeded in demonstrating the ease with which the so-called organic molecules are created abiotically. Miller's (1953) classical experiment sparking ammonia and hydrogen gas above water in a sealed apparatus produced amino acids and sugars, common building blocks of life as we know it on Earth. These experiments have since been repeated with a variety of gas mixtures, with lower but still significant yields of organic matter in CO_2-rich mixtures (Schopf, 1983). Yields are promoted by addition of clays and chemically reduced minerals such as pyrite and siderite (Schoonen et al., 2004). Clays also promote assembly of oligomers of RNA up to 50 units long (Ertem and Ferris, 1996). Clays act as catalysts, templates, compartments, energy storage devices, and chemostats in promoting organic molecule synthesis. Siderite can mediate photooxidation of CO_2 to produce reduced carbon in a way comparable with photosynthesis (Braterman et al., 1983). Concentrated amino acids (proteinoids) can form spherical structures (Fox and Dose, 1972), which appear superficially like living cells, but are far from functioning cells (**Figure 3**). Nothing close to a living cell has ever emerged from experiments designed to examine the origin of life, so such experimental research into the origin of life has reinforced the enormous gulf between quick organic matter and dead organic matter.

9.10.4.2 Geologic Records

Carbonaceous chondrites, with their mix of smectite clay and diversity of reduced organic compounds (Pizzarello et al., 2001), are evidence for abiotic synthesis of organic compounds in the presence of clay and reduced iron minerals

such as siderite early during the origin of the solar system. The carbonaceous clayey matrix postdates olivine chrondules that are 4.56 Ga in age (Amelin et al., 2002) and predates cross-cutting veins of calcite (Endress et al., 1996) and halite (Whitby et al., 2000) no younger than 4.51 Ga. The characteristic carbon isotopic fraction of rubisco, an important enzyme in photosynthesis, is indicated by light carbon isotopic ratios of amorphous organic matter in the 3.8 Ga Isua greenstone belt of west Greenland (Rosing and Frei, 2004), and microfossils comparable with cyanobacteria are found in chert from the 3.5 Ga Apex Chert of the Warrawoona Group near Marble Bar, Western Australia (Schopf, 1983). Thus, the origin of life occurred during the 700 million year interval between 4.5 and 3.8 Ma or approximately the first 15% of the history of our planet. This period is a dark age in Earth history, its sedimentary records obliterated by metamorphism and subduction. Surfaces of the Moon and Mars dating to this early time in the history of the solar system also reveal another reason for paucity of geologic records was heavy bombardment of larger planets by planetesimals and other debris of the evolving solar system. One of these very large impacts at about 4.4 Ga is thought to have created our Moon from a Mars-sized impactor ('Theia') melting itself and a large fraction of the Earth (Halliday, 2004). Impact melting of rocks would have destroyed both life and any precursor clay or organic matter (Maher and Stevenson, 1988). Another impedance to life and its precursors in surface environments would have been ultraviolet and other forms of protein-denaturing radiation from a young sun unfiltered by a fully formed atmosphere and ozone layer (Sagan and Pollack, 1974).

9.10.4.3 Soil

The idea that life came from soil is probably the most ancient human view of the origin of life. A cuneiform text dating to 2000 BC from the Sumerian city of Nippur in present-day Iraq (Kramer, 1944) describes a feast of the gods, presided over by Ninmah (mother earth) and Enki (god of the waters), in which Ninmah created different kinds of humans from clay and Enki decreed their role in life. The idea of the spontaneous generation of life from soil was popular for centuries, endorsed by such influential thinkers as Aristotle (c.384–322 BC) and Lucretius (c.99–55 BC), but lost much appeal in the late nineteenth century when Louis Pasteur showed that life did not arise spontaneously from the soil or organic matter, but from microscopic propagules. Nevertheless, soil has much promise as a site for the origin of life because of its self-organizing complexity, lack of chemical equilibrium, and benign physicochemical conditions (Retallack, 2001a). The soil is alternately full and then drained of water after rainstorms, but thin films of water persist in pockets and margins of grains in which complex chemical reactions stop and start with the influx and evaporation of soil water (**Figure 7**). Weathering is a slow process taking millions of years to reach chemical equilibrium in the production of nutrient-starved kaolinitic clays. Soils are full of insulating clays, which moderate heat differences; diaphanous grains, which filter harsh electromagnetic radiation; and base-rich clays, which neutralize environmental acids.

9.10.4.4 Sea

The idea that "life like Aphrodite was born on the salt sea foam" (Bernal, 1967) has a long pedigree extending back to Ionian Greek philosophers Thales (c.585 BC) and Anaximander (c.565 BC). The appeal of this idea comes partly from the abundance of life in the sea at present and the aqueous geochemistry of most vital processes. There is also the saline composition of blood plasma and cytoplasm of land animals, suggestive of marine origins. Darwin's (1959) posthumously published vision of life's origins in "some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc." hints at the theory's the greatest weakness: the remarkably uniform and dilute composition of the ocean, which tends toward chemical and physical equilibrium (Nisbet and Sleep, 2001). The smaller the pond and the more intermixed with mineral matter, the better opportunity for complex serial biosynthesis.

9.10.4.5 Deep-Sea Vent

Ocean floor volcanic vents discovered by deep-ocean submersibles are yet another plausible site for the origin of life (Corliss et al., 1981). Hot water billows out of these vents like black smoke because of the rapid precipitation of sulfides and other dissolved substances into cold seawater. The vent fluids are like toxic waste, highly acidic, oxygen-starved, and extremely hot (temperatures up to 380 °C). There is little organic carbon or organisms immediately around them, but inches away the sea floor is crowded with life: large white clams and crabs and peculiar pale tube worms. Only a few meters beyond, cold temperatures and lack of nutrients severely curtail biodiversity. At such great depths, there is no possibility of photosynthesis, so the community is fed by chemosynthetic microbes (**Table 2**). Life is envisaged to have originated in spongy rock and vent deposits between the hot, caustic vents and cold, dilute ocean. Hydrothermal vent origin of life is an appealing explanation for widespread heat-shock proteins and primitive thermophilic bacteria (Nisbet and Sleep, 2001), though it is debatable whether these adaptations are primitive (Forterre, 1995). Unlike soils or tidal flats, deep-sea vents are not controlled in their location or persistence by microbial scums. Black smokers erupt and become dormant like volcanoes, by virtue of subsurface faulting creating conduits for fluid migration in their oceanic rift valleys.

9.10.4.6 Other Worlds

Another possibility is that life evolved elsewhere in the universe and colonized our planet as propagules that could withstand long-distance transport in space within rocks or resistant pellicles. This concept of 'panspermia' goes back to the turn of the century and the Swedish chemist Svanté Arrhenius. A related idea is deliberate colonization of the Earth by advanced extraterrestrial civilizations (Crick, 1981). All manner of organisms could have been broadcast or dispatched, ranging from influenza viruses or unicellular bacteria to sophisticated aliens in space vehicles. Such views have some appeal in this age of space exploration but merely remove the question of the origin of life to another planet. The

environment where that life evolved is likely to have been Earth-like in many respects because life has long been well suited to our planet. Thus, it remains useful to consider the origin of life from natural causes here on Earth.

9.10.5 Coevolutionary Histories

Biosignatures of Earth can be understood by their long geologic histories, which present natural experiments with variation that may reveal oscillatory mechanisms of coevolution. The metaphor of an arms race is commonly applied to coevolutionary trajectories because they appear to change toward more competitive organisms (Dawkins and Krebs, 1979). Increasingly discriminating bees are more faithful to the pollination of particular flowering species. Grassy sod with basal-tillering and telescoped internodes is best able to absorb the rough grazing of tall-toothed and hard-hooved bison and horse. Similarly in warfare, arms races between competing nations result in more sophisticated and powerful weapons. Unlike biological coevolution, which has spawned diversity, arms races commonly result in imperialism and loss of other cultures, their languages, and archives.

Another metaphor for coevolution is the Red Queen hypothesis (Van Valen, 1973) named for Lewis Carroll's character who had to run faster and faster in order to stay in the same place, presumably within a social pecking order. This implies that all creatures are caught up in a rat race and does not account for the persistence of such slow creatures as millipedes and possums.

I prefer a more fundamental human metaphor of technological escalation, which applies also to activities such as information technology (Vermeij, 2004). Hand copying, printing, and electronic publication are successive improvements in information technology driven by inventions (adaptations) of companies (producers) to satisfy demands of customers (consumers), who in turn adjust their consumption (coadaptations). Technological escalation in information technology has resulted in not only increasing quality and volume of communication between groups but also increasing diversity and specialization.

By any of these models, the history of life and Earth environments should demonstrate a pattern of adaptations in one segment of the ecosystem followed by compensatory coadaptations in another part of the ecosystem. The following sections attempt to identify such reciprocal changes in air, water, soils, and rocks. There are also catastrophic alterations of surface environments after large bolide impacts, giant volcanic eruptions, and metamorphism of limestone and coal (Retallack, 2002). There are biological perturbations, such as seasonal leaf-shedding and soil fertility cycles, on such short timescales (10^0–10^5 years) as well (Retallack, 2004a). Biological flexibility on these timescales may have aided biotic recovery from abiotic catastrophes. The present account, however, will deal only with evolutionary timescales (10^6–10^9 years).

9.10.5.1 Life and Air

Bioterraforming, or reengineering of planetary environments by life, as may be possible for Mars (Fogg, 1998), could already

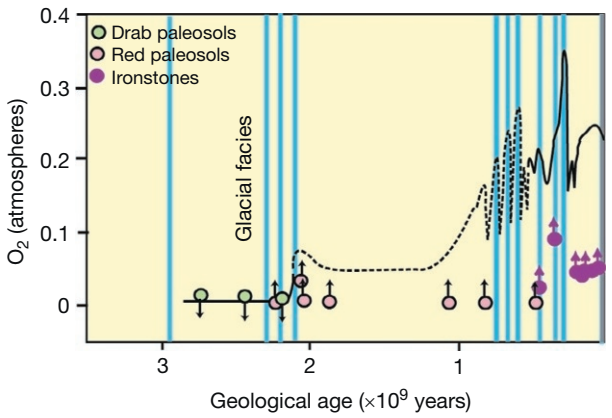


Figure 8 Reconstructed history of atmospheric oxygen abundance, showing constraints from selected Precambrian paleosols and Phanerozoic pedogenic goethites (updated from data of Retallack, 2001a). The Phanerozoic curve (solid line) is from a sediment mass balance computer model (Berner et al., 2000) and the Neoproterozoic curve (dashed line) inferred from carbon isotopic data (Anbar and Knoll, 2002). The gray bars represent times of known glacial sedimentation, which vary considerably in duration and recurrence interval.

have happened on Earth during the Precambrian evolution of cyanobacterial photosynthesis to create our oxygen-rich atmosphere (Kirschvink and Kopp, 2008). Evidence from paleosol geochemistry suggests an especially marked oxidation at about 2.3 Ga (**Figure 8**), as does the coeval demise of banded iron formations (Bekker et al., 2004), fluvial uraninite placers (England et al., 2002), and mass-independent fractionation of sulfur isotopes (Farquhar et al., 2000). Difficulties for this scenario of possible geologically older cyanobacterial microfossils (3.5 Ga; Schopf, 1983), stromatolites (3.5 Ga; Allwood et al., 2006), mats (3.4 Ga; Tice and Lowe, 2004), and 2α-methylhopane biomarkers (2.7 Ga; Brocks et al., 2005) have now been falsified by Kirschvink and Kopp (2008): neither the biomarkers nor the megascopic and microscopic forms are unique to cyanobacteria.

A more recent atmospheric oxidation event is the peak value of 35% O_2 during 390 Ma spread of forests, followed by a 250 Ma low of about 15%, before climbing again to the modern value of 21% with the 19 Ma spread of sod grasslands (**Figure 8**). Forest ecosystems of the Middle Devonian (390 Ma) culminated a steady increase in woody plant size from the first appearance of matchstick-sized primitive land plants during the early Silurian (430 Ma). These evolutionary changes were matched by comparable increases in depth and intensity of weathering indicated by paleosols, suggesting that not only individuals evolved but also soil communities distributed over broad continental areas (Retallack, 1997a). Woody steles of the earliest land plants frustrated early Silurian herbivores, which were mainly millipedes adapted to feeding on unvascularized liverworts (Retallack, 2001c). A coevolutionary swing toward woody plants to the disadvantage of herbivorous animals would not be enough in itself to change the world, unless such communities became globally distributed, commandeering the source of mineral nutrients in soil, metabolic gases in the atmosphere, and the distributive capacity of rivers and the ocean. The evolution of laminar

leaves by 320 Ma would have enhanced photosynthetic production of wood and transpiration of water (Beerling et al., 2001). Altering atmospheric composition requires not only living communities as a force for photosynthetic oxidation but also deep and protracted silicate weathering and burial of carbon in swamps and oceans in large quantities (Berner et al., 2000).

Silicate weathering and carbon burial could be regarded as physicochemical components of global change, but both are under strong biotic control. The simple inorganic weathering agent, carbonic acid, from dissolution of CO_2 in water comes largely from soil respiration, in turn fed by primary production, so that some rainforest soils have 110 times more soil CO_2 than in the atmosphere (Brook et al., 1983). Other organic acids also enhance hydrolytic weathering (Berner, 1997) and are also more abundant in larger and more productive ecosystems. The physical evidences of this biotic enhancement of weathering are the new kinds of forest soils (Alfisols and Ultisols), which appear as paleosols in Devonian and younger rocks (Retallack, 1997a). Carbon burial can be regarded as a purely physical process, but the biological invention and then proliferation of lignin first occurred in a world without effective ligninase enzymes or creatures such as termites and dinosaurs that could reduce the load of woody debris undecayed and finding its way into rivers, lakes, swamps, and the sea. The physical evidence of Devonian–Carboniferous increase in carbon burial is not only the increase in marine and lacustrine black shales but also the appearance of a new kind of peaty soil (Histosols) and wetland-forest (swamp) ecosystem in Devonian and younger rocks. Thus, the soaring oxygen content of 35% was largely a product of the balance between producers with the newly evolved product of wood getting ahead of consumers. By Late Carboniferous, there were few vertebrate herbivores but many winged herbivorous insects, wood-eating cockroaches, and fungi. Carboniferous swamp woods were punky with soft cells (parenchyma) between the files of lignin-rich cells (tracheids) and so were less prone to wildfire than many modern woods. Carboniferous ecosystems thus included features that offset the oxidizing effects of trees, culminating in the evolution of tree-eating termites and dinosaurs during the long Mesozoic CO_2 greenhouse (Retallack, 2004a).

Sod grasslands formed by a coevolutionary process between grasses and grazers (**Figure 2**) on all continents except Australia and Antarctica over a geologically significant interval of time between the late Eocene (35 Ma) and Pliocene (4 Ma). Studies of root traces in paleosols indicate that before grasslands, the stature of woody vegetation became uniformly smaller from humid climate forests to arid shrublands, as it still does in Australia where mallee vegetation of 2–10 m multistemmed shrubs dominates a semiarid climatic belt of grasslands on other continents (Retallack, 2012). From late Eocene to late Miocene, grasslands were confined to the rainfall belt of 300–400 mm mean annual precipitation, but after Pliocene evolution of large hard-hooved hypergrazers such as horses and wildebeest and megaherbivores such as elephants, the prairie-forest ecotone was rolled back nearer to the 1000 mm isohyet of mean annual precipitation (Retallack, 2001b). Evidence for this territorial advance of grasslands comes from the spread of the characteristic surface horizons of grasslands (mollic epipedon) in soils with deep calcareous nodular horizons (calcic horizon) of subhumid climates. Human coevolution with grass-based agroecosystems has now created grasslands even in rainforest climate belts. This territorial expansion is evidence that this coevolved ecosystem is not merely an assemblage of creatures with mutually compatible climatic tolerances, but a biological force for global change.

Like forests of the Carboniferous, Cenozoic grasslands were a force for atmospheric oxidation because they increased plant biomass, particularly underground, compared with mallee vegetation in comparable climatic belts. Grasslands also promoted hydrolytic weathering, accelerating its pace with a rich soil fauna of earthworms and a variety of microbes. The result is a dramatic increase in soil organic carbon contents as high as 10 wt% for depths of up to a meter, compared with mallee soils of comparable climatic belts that have such high organic content only in the surficial few centimeters of leaf litter. Soil organic matter also stabilizes a characteristic soil structure of crumbs, which withstands wetting and can be transported considerable distances by streams to be buried in agricultural stock ponds, lakes, river floodplains, and the ocean. Thus, Cenozoic grasslands increased efficiency of biomass creation, hydrolytic weathering, and carbon burial in arid lands, just as forests did during their Devonian–Carboniferous evolution in humid areas and wetlands. Grasslands also effected global change by controlling two additional landscape attributes traditionally regarded as purely physical: albedo and evaporation (Retallack, 2001b). Grasses are lighter-colored than trees, especially when hay-colored during dry or snowy seasons, and bow down under snow more readily than trees. Grasslands thus reflect more radiation back into space, with a net chilling effect on the planet. Grass colonization of the bare soil patches of desert shrubland or mallee has the effect of reducing evaporation of soil water. Grassy sod is moist and the air above grasslands dry, in contrast to woodlands and forests that have dry soil and moist air because of persistent transpiration of water by trees. Lowered vapor pressure of water over grasslands compared with forests also has the effect of planetary cooling, because water vapor is a powerful greenhouse gas. Thus, the rise of O_2 to 21% with Cenozoic drying and cooling culminating in the Pleistocene ice age may have been in part due to the coevolution of grasslands. The process is not without limits, but even the warming forces of greenhouse gas (CO_2 and CH_4) emissions from wildfires and large herds of ungulates are part of a system that promotes grassland expansion at the expense of forests, because grasses recover from fire and herbivory more readily than trees. If anything can undo the cooling effect of grasslands, it would have to be their domineering evolutionary product, us (Palumbi, 2001).

9.10.5.2 Life and Water

Ocean composition also has changed on evolutionary timescales. Oceans of the early Earth may have been 1.5–2 times more saline than modern oceans, judging from fluid inclusion and evaporite volume data (Knauth, 1998). The principal mechanism for oceanic dilution is burial of large volumes of salt in evaporite deposits of barred basins, like the modern Persian Gulf. The growth of continental granitic crust promoted the tectonic formation of such basins and emergent

continental storage of evaporites. Barred evaporite basins also are common behind coral, algal, and stromatolitic reefs, which extend back at least to the Paleoproterozoic (2 Ga; Grotzinger, 1988). Paleoproterozoic was also a time of peak abundance of banded iron formations (Figure 9), the distinctive laminated hematite and chert deposits of oceans during the transition from chemically reducing to chemically oxidizing oceans due to the spread of photosynthetic microbes (Cloud, 1976).

A role for life in Precambrian changes in salinity and redox remains controversial because they are so remote in geologic time. A clearer indication of life's role in oceanic chemistry on evolutionary timescales is the Phanerozoic oscillation between aragonite and calcite seas (Figure 10). This geochemical oscillation is seen in both carbonates and evaporites of the Cambrian to Mississippian (540–320 Ma) and Jurassic to Paleogene (170–20 Ma), which are dominantly calcite and sylvite–gypsum. In contrast, carbonates and evaporites of the Mississippian to Jurassic (320–170 Ma) and Neogene (20–0 Ma) are dominantly aragonite and kieserite–gypsum. The salts are simple precipitates, and although the carbonates are largely precipitated as skeletons of marine organisms, the major rock-forming organisms use passive forms of extracellular metabolically induced precipitation, rather than shell formation mediated by an organic matrix (Stanley and Hardie, 1999). Calcite seas correspond broadly in time with high-CO_2 greenhouse intervals and aragonite seas with low CO_2 icehouse intervals, but this does not appear to be a simple pH effect from a more acidic atmosphere at times of high CO_2 and carbonic acid. Calcite is favored over aragonite only for a narrow range of unrealistically high partial pressures of CO_2. Instead, calcite versus aragonite is favored by the Mg/Ca ratio of the ocean, with calcite favored when the ratio is less than 2 and aragonite favored when it is more than 2. The Mg/Ca ratio of the ocean is in turn controlled by the balance of cations in solution from rivers and from hydrothermal activity at mid-ocean ridges. Stanley and Hardie (1999) argued that variation in mid-ocean ridge spreading is the primary control, with active spreading inducing a greater hydrothermal flux and subseafloor sequestration of Mg. However, rates of sea floor spreading are not appreciably varied over the past 180 Ma in which spreading can be inferred accurately from seafloor magnetic striping (Rowley, 2002). An alternative control on oceanic Mg/Ca ratio is riverine input, considering the greater abundance of dolomitic caliche in paleosols of high-CO_2 greenhouse climates than those of low-CO_2 icehouse climates (Retallack, 2004a). Dolomitic versus calcitic caliche in turn is probably under control of different microbial polysaccharides (Zhang et al., 2012). From their source as they are weathered from soils to their sink in fossil coral reefs, Ca and Mg cations are under constant biotic surveillance.

9.10.5.3 Life and Soil

Compared with a long history of research on Precambrian air and water, even such basic questions as when soils became living remain uncertain. There was land and thick deeply weathered soil at least as far back as 3.5 Ga, as indicated by paleosols of that age (Buick et al., 1995), and isotopic geochemical evidence for substantial amounts of archaean continental crust (Bowring and Housh, 1995). Sterilization of land

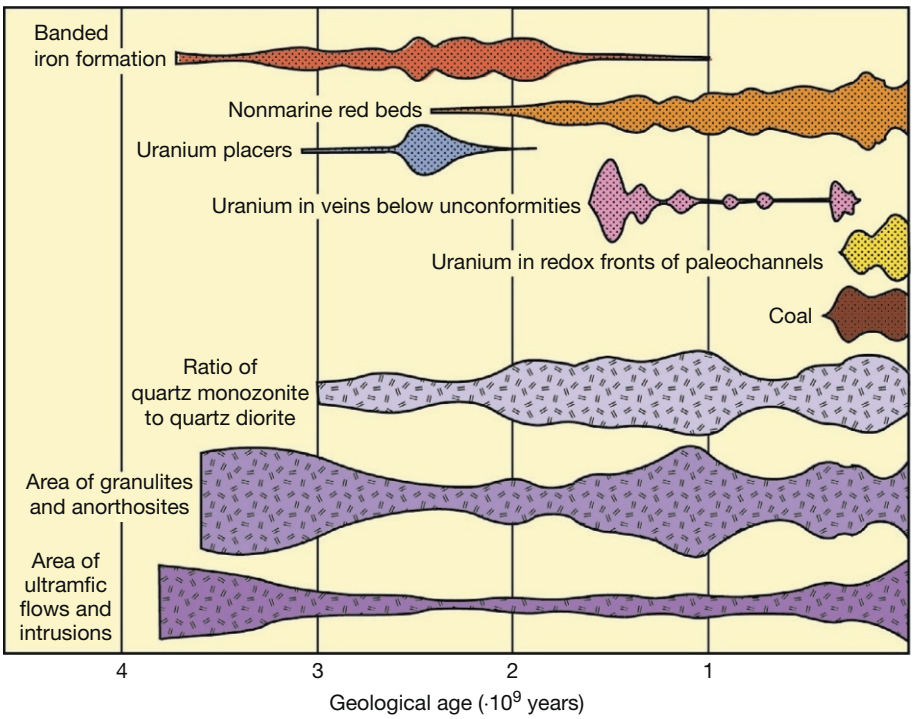


Figure 9 Relative abundance of sedimentary rocks, ores, and igneous rocks through geologic time. Adapted from Engel AE, Itson AP, Engel CG, Stickney DM, and Cray EJ (1974) Crustal evolution and global tectonics: A petrogenic view. *Geological Society of America Bulletin* 85: 843–858; Meyer CE (1985) Ore metals through geologic history. *Science* 227: 1421–1428; Barley ME and Groves DI (1992) Supercontinent cycles and the distribution of metal deposits through time. *Geology* 20: 291–294.

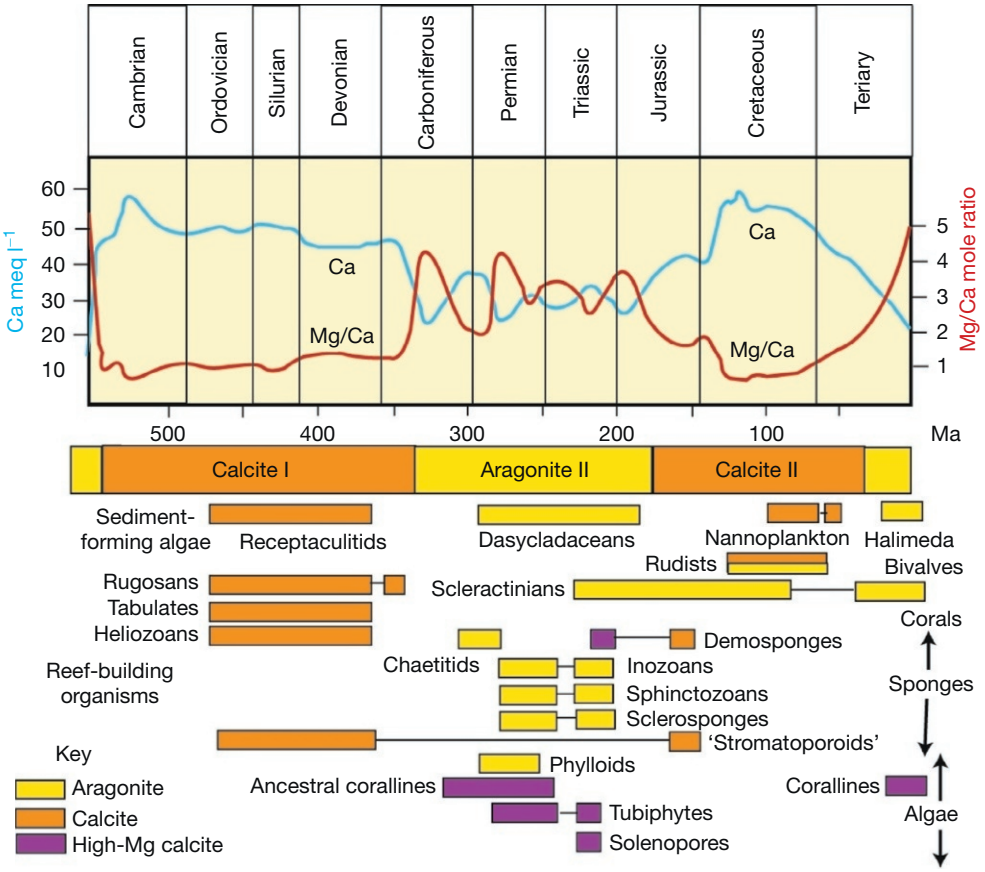


Figure 10 Correspondence between oceanic alkaline earth composition and the carbonate mineralogy of hypercalcifying marine organisms. Simplified from Stanley SM and Hardie LA (1999) Hypercalcification: Paleontology links plate tectonics and geochemistry to sedimentology. *GSA Today* 9(2): 1–7.

surfaces by intense ultraviolet radiation before development of the ozone layer is unlikely in an early atmosphere of methane, water vapor, or other dense early atmospheric gases (Rye and Holland, 1998). The very existence of clayey paleosols that were well drained is regarded by some geomorphologists as evidence of stabilization by life, including microbes (Schumm, 1977). Remarkably light C-isotope values (−40 δ^{13}C‰ vs. Pee Dee Belemnite (PDB)) indicative of methanotrophs have been recorded from the 2.8 Ga Mt Roe paleosol of Western Australia (Rye and Holland, 2000) and isotopically heavy C-isotope values (−16 to −14 δ^{13}C‰) like those of hypersaline microbes from the 2.6 Ga Schagen paleosol of South Africa (Watanabe et al., 2000). Many paleosols have carbon isotopic values in the normal range for cyanobacteria and other photosynthetic organisms (−24 to −30 δ^{13}C‰): 2.3 Ga pre-Huronian in Ontario (Farrow and Mossman, 1988), 1.3 Ga post-Mescal paleokarst in Arizona (Beeunas and Knauth, 1985), and 0.8 Ga pre-Torridonian in Scotland (Retallack and Mindszenty, 1994). To this list of methanogenic, hypersaline and photosynthetic life in Precambrian paleosols must be added decomposers such as actinobacteria, because most Precambrian paleosols have only traces of organic C. Something must have cleaned up the dead bodies, just as fungi and other microbes create carbon-lean modern soils. Filamentous cyanobacterial microfossils have been found permineralized on a paleokarst paleosol in the 1.3 Ga Mescal Limestone of Arizona (Horodyski and Knauth, 1994) and within angular flakes of laminated crust in the 2.8 Ga Mt Roe paleosol of Western Australia (Rye and Holland, 2000). Comparable modern microbes have wide environmental tolerances, and comparable fossils are common in permineralized stromatolites (Schopf, 1983), so were they soil microbiota or just opportunistic colonists of puddles? Lichen-like microfossils from the Carbon Leader of the Witwatersrand Group of South Africa (Hallbauer et al., 1977), now known to be 2.9 Ga in age (England et al., 2002), have been discounted as artifacts of acid maceration (Cloud, 1976), although it is clear from petrographic thin sections that they predate metamorphic veins (MacRae, 1999) and were redeposited during sedimentation (Mossman et al., 2008). The organic matter of these structures has isotopically light carbon, pristane, phytane, and pentose≈hexose indicative of photosynthesis (Prashnowsky and Schidlowski, 1967). Life in soil may go back 3.5 Ga and perhaps further (**Figure 7**).

Soils played a role in coevolutionary oscillations of atmospheric and oceanic chemical composition (**Figures 8–10**). One of the best lines of evidence for a chemically reducing early atmosphere on the early Earth is paleosols older than 2.2 Ga with the deep spheroidal weathering suggestive of a well-drained soil yet chemically reduced clay and iron minerals (**Figure 4(a)**). These appear to be a completely extinct form of

soil, which I have labeled 'Green Clays' (Retallack, 2001a). The degree and depth of weathering of paleosols on the early Earth are remarkable, with some of them attaining the alumina enrichment of bauxites (Figure 11). The appearance of laterites and oxidized paleosols is one of the most obvious lines of evidence for Paleoproterozoic atmospheric oxidation (Rye and Holland, 1998). The oldest known caliche may be the Schagen paleosol (2.6 Ga) of South Africa (Watanabe et al., 2000). The first silcrete appears along with the first desert dunes and sizeable (>600 km) continents in the pre-Pitz Formation paleosols (1.8 Ma) of northern Canada (Ross and Chiarenzelli, 1984). Silcretes and caliche are characteristic of deserts on Earth today, and it is surprising not to find them on the earlier Earth. Primordial rolling terrains of Green Clay paleosols would have seemed unusual by modern standards. Without the benefit of multicellular vegetation, these landscapes would have appeared barren and desolate, and yet these deep clayey soils lacked the harsh rocky outlines and dunes of modern desert landscapes. Their deep weathering can be attributed to a humid, maritime climate and warm temperature assured by the greenhouse effect of elevated levels of atmospheric water vapor, methane, and carbon dioxide (Figure 12).

The role of soil in Phanerozoic environmental oscillations is especially clear from trace and body fossils associated with paleosols. Burrows and tracks of millipedes first appear in mid-Ordovician (460 Ma) paleosols (Retallack, 2001c, 2011) and termite nests in Triassic (230 Ma) paleosols (Hasiotis and Dubiel, 1995). These animals and their burrows had profound effects on soil aeration and respiration. The effect of plants is even more obvious, with reduction spotting and rhizome traces first appearing in early Silurian (430 Ma) paleosols and large root traces of trees in Middle Devonian (390 Ma) paleosols (Retallack, 1997a; Retallack and Huang, 2011). The evolution of trees coincided with evolution of forest soils (Alfisols) and was soon after followed by forest adaptations to low-nutrient conditions of mineral-poor peats (Histosols) by Late Devonian (365 Ma), silica sands by Mississippian (330 Ma), and base-poor clays (Ultisols) by Pennsylvanian (320 Ma; Retallack, 2001a). The evolution of sod grasslands with their characteristic surface horizons (mollic epipedon) ushered in the first grassland soils (Mollisols) in the Early Miocene (19 Ma; Retallack, 2001b). The oscillation of terrestrial communities dominated by millipedes, forests, termites, and grasslands corresponds to grand swings in greenhouse to icehouse paleoclimates (Retallack, 2004a).

9.10.5.4 Life and Rocks

The abundance of different kinds of rocks has also waxed and waned with life-mediated paleoenvironmental changes on Earth (Figure 9). Uranium ores, for example, show three distinctive forms, which succeed one another over geologic history. Uranium in fluvial sandstones is found as rounded grains that were evidently transported by rivers in well-aerated turbulent water and trapped in less turbulent portions of the stream with gold and other heavy grains in a kind of deposit called a placer (England et al., 2002). These dense grains of uraninite are highly unstable in modern rivers because they readily oxidize to soluble yellow carnotite. Evidence from chemically reduced but deeply weathered paleosols beneath these uraniferous sandstones supports the notion that uraninite placer deposits reflect a distant time (1.9–3.1 Ga) of much lower atmospheric oxygenation (Mossman et al., 2008). In geologically younger rocks (0.2–2.0 Ga), uraninite ores fill veins and breccias associated with unconformities, where they precipitated at a zone of chemical reduction after transportation in overlying sandy aquifers from granitic sources as oxides in surface and groundwaters. Geologically younger again (0.3–0 Ma) are roll-type uranium ores, which reflect similar oxidized transportation in groundwater until encountering a redox front in fluvial sandstones locally rich in organic matter such as fossil logs (Meyer, 1985). Thus, uranium ore types are related to atmosphere and aquifer redox, which in turn are biologically mediated (Retallack, 2004a).

Common kinds of igneous rocks also show variations in relative abundance that may reflect biologically mediated paleoenvironmental changes (Figure 9). The early Earth like

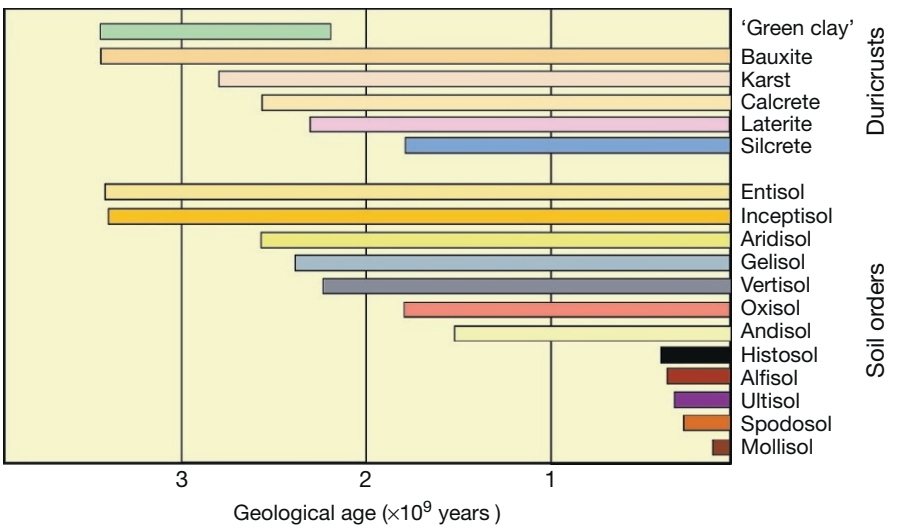


Figure 11 Stratigraphic range of duricrusts and soil orders. Entisols and Inceptisols are assumed to have been precursors of other paleosols, even though difficult to recognize in the rock record before the advent of roots and burrows in soils (updated from data of Retallack, 2001a).

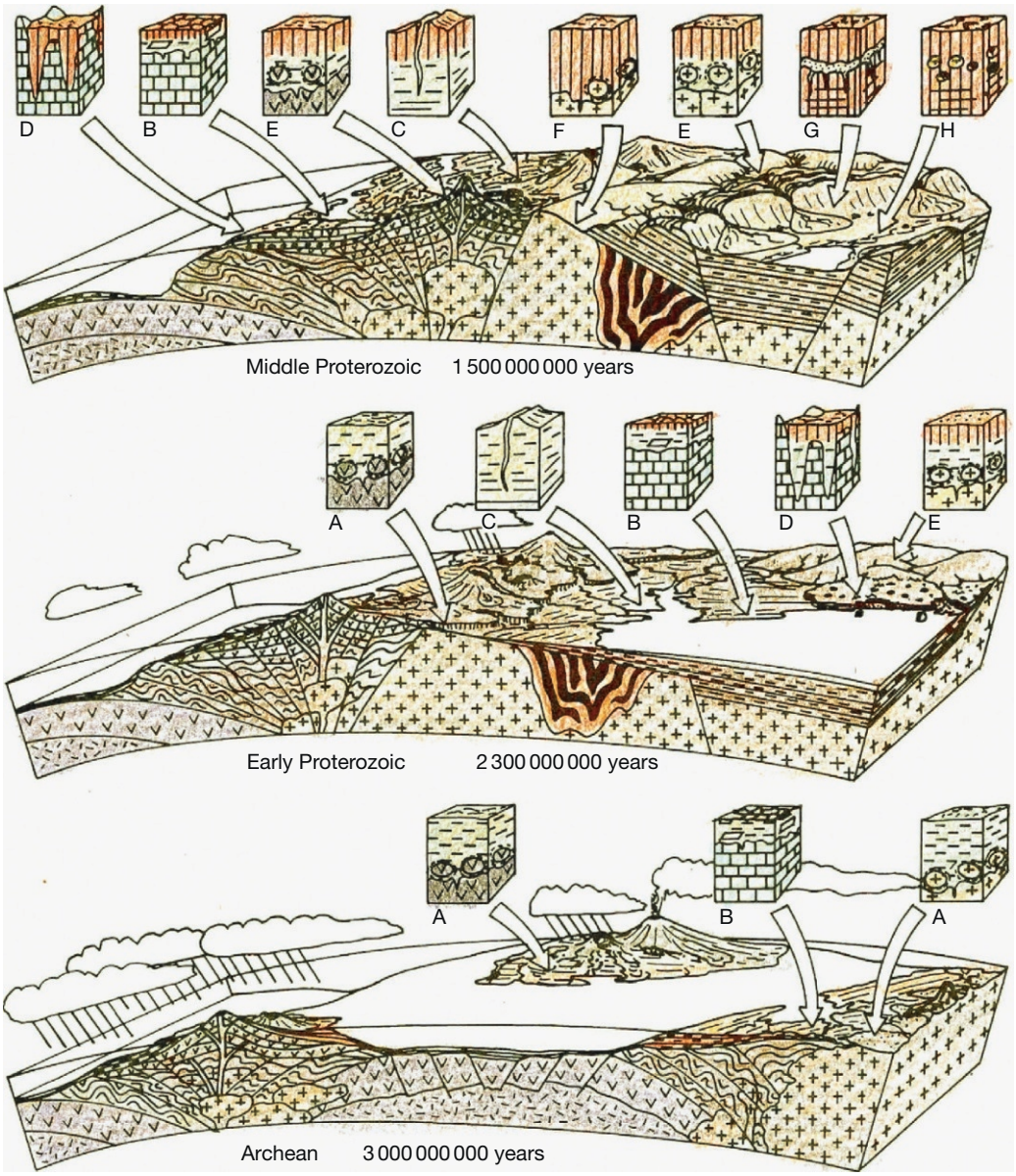


Figure 12 A speculative scenario for the evolution of soils, atmosphere, and continents on the early Earth, in which desert soils appear long after deeply weathered soils. The soil types illustrated are extinct 'Green Clays' (A), salty soils or Salids (B), swelling clay soils or Vertisols (C), karst and drab cave earth or Orthents (D), oxidized incipient soils or Ochrepts (E), red and deeply weathered soils or Oxisols (F), desert soils with silcretes or Durids (G), and desert soils with calcareous horizons or Calcids (H). Reproduced from Retallack GJ (2001) *Soils of the Past*, 2nd edn., p. 600. Oxford: Blackwell, with permission.

the Moon, Venus, and Mars was dominated by basaltic and ultramafic igneous rocks. The proportion of quartz-rich igneous rocks such as quartz monzonite and granite to diorite and other more mafic rocks increase steadily after 3 Ga, reaching peaks at about 1.8, 0.8, 0.4, and 0.2 Ga. The latter two corresponding to Ordovician and Jurassic are the familiar CO_2 greenhouse (**Figure 8**), calcite-sea (**Figure 10**), and animal-dominated soil episodes (Retallack, 2004a). These were times of widespread deep weathering in a warm–wet world lacking ice caps.

Other rocks may also be under surficial environmental control. For example, eruptions of basaltic andesites have been related to surface conditions, because volcanoes of northern California preferentially erupt during isostatic rebound after unloading of melted ice caps and sea level rise (Jellinek et al., 2004).

9.10.6 Conclusions

A coevolutionary perspective explains many aspects of our Earth's distinctive geologic history. As in classical cases of coevolution, such as grasses and grazers, interactions between unrelated but mutually dependent organisms promote the persistence of new and earth-changing ecosystems, such as grasslands (Retallack, 2001a). The global change capabilities

of such ecosystems arise because organisms are evolving primarily in response to other organisms and only indirectly in response to their physicochemical environment. As coevolved ecosystems proliferated to commandeer the nutrient supply of soil, the metabolic gases of the air, and the medium of water, they altered the atmosphere, hydrosphere, pedosphere, and deep lithosphere.

At the heart of any coevolutionary process is the natural selection of specific coadaptations to other components of the biota. These inventions are rare events that appear to flaunt the laws of thermodynamics and entropy, because they promote continuing metabolism and disorder (Vermeij, 2004). As rare events, such adaptations are followed at long geologic intervals by coadaptations, with the directed trends of coevolution achieved through a series of discrete oscillations. As these ecosystems spread and proliferated, coevolutionary oscillations of life were transferred to the geologic history of our air, water, soil, and rocks (Retallack, 2004a).

From this perspective, human agroecosystems are not the first coevolved ecosystem to induce global warming. Just as cyanobacterial mats cooled a world of methanogenic slimes about 2 Ga, forests cooled a world of millipedes about 390 Ma, and grasslands cooled a world of large mammals about 35 Ma, we can hope that new coevolutionary initiatives will restore Earth to a livable temperature. If we fail in this mission, other organisms may succeed.

Acknowledgments

I thank Geerat Vermeij, Alberto Patiño-Douce, Peter Kelemen, Bill Schopf, Bruce Runnegar, Peter Ward, and Alex McBirney for their philosophical and revealing discussions and the University of Oregon for unusual academic freedom over more than two decades.

References

Allwood AC, Walter MR, Kamber BS, Marshall CP, and Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* 441: 714–718.

Amelin Y, Krot AN, Hutcheon ID, and Ulyanov AA (2002) Lead isotopic ages of chondrules and calcium–aluminum-rich inclusions. *Science* 297: 1678–1683.

Anbar AD and Knoll AH (2002) Proterozoic ocean chemistry and evolution: A bioinorganic bridge. *Science* 297: 1137–1141.

Bada JL, Glavin DP, McDonald GD, and Becker L (1996) A search for endogenous amino acids in Martian meteorite ALH84001. *Science* 279: 362–365.

Bandfield JL, Glotch TD, and Christensen PR (2003) Spectroscopic identification of carbonate minerals in Martian dust. *Science* 301: 1084–1087.

Barley ME and Groves DI (1992) Supercontinent cycles and the distribution of metal deposits through time. *Geology* 20: 291–294.

Barsukov VL, Volkov VP, and Khodakovsky IL (1982) The crust of Venus: Theoretical models of chemical and mineral composition. *Journal of Geophysical Research* 87A(supplement): 3–9.

Beerling DJ, Osborne CP, and Chaloner WG (2001) Evolution of leaf form in land plants linked to CO_2 decline in the late Palaeozoic era. *Nature* 410: 352–354.

Beeunas MA and Knauth LP (1985) Preserved stable isotopic composition of subaerial diagenesis in the 1.2 b.yr Mescal Limestone, central Arizona: Implications for the timing and development of a terrestrial plant cover. *Geological Society of America Bulletin* 96: 737–745.

Bekker A, Holland HD, Wang P-L, et al. (2004) Dating the rise of atmospheric oxygen. *Nature* 427: 117–120.

Bernal JD (1967) *The Origin of Life*, p. 345. New York: World,.

Berner RA (1997) The rise of plants and their effects on weathering and CO_2. *Science* 276: 543–546.

Berner RA, Petsch ST, Lake JA, et al. (2000) Isotope fractionation and atmospheric oxygen: Implications for Phanerozoic O_2 evolution. *Science* 287: 1630–1633.

Bonnefoi CC, Provost A, and Albarèdo F (1995) The "Daly gap" as a magmatic catastrophe. *Nature* 378: 270–272.

Borg L, Connelly JN, Nyquist LE, Shih C-Y, Wiesman H, and Reese Y (1999) The age of carbonates in Martian meteorite ALH84001. *Science* 286: 90–94.

Bowen NL (1922) The reaction principle in petrogenesis. *Journal of Geology* 30: 177–198.

Bowring SA and Housh T (1995) The Earth's early evolution. *Science* 269: 1535–1540.

Brasier MD, Green OR, Jephcoat AP, et al. (2002) Questioning the evidence for Earth's oldest fossils. *Nature* 419: 76–80.

Braterman PS, Cairns-Smith AG, and Sloper RW (1983) Photooxidation of hydrated Fe^{2+} enhanced by deprotonation: Significance for banded iron formations. *Nature* 303: 163–164.

Brearley AJ (1997) Disordered biopyriboles, amphiboles and talc in Allende meteorite: Products of nebular or parent body alteration? *Science* 276: 1103–1105.

Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, and Bowden SA (2005) Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* 437: 866–870.

Brook GA, Folkoff ME, and Box EO (1983) A world model of soil carbon dioxide. *Earth Surface Processes and Landforms* 8: 79–88.

Buick R, Thornett JR, McNaughton NJ, Smith JB, Barley ME, and Savage M (1995) Record of emergent continental crust approximately 3.5 billion years ago in the Pilbara Craton of Australia. *Nature* 375: 574–577.

Bunch TE (1975) Petrography and petrology of basaltic achondritic breccias (howardites). *Geochimica et Cosmochimica Acta* 6: 469–492.

Bunch TE and Chang S (1980) Carbonaceous chondrites II. Carbonaceous chondrite phyllosilicates and light element geochemistry as indicators of parent body processes and surface conditions. *Geochimica et Cosmochimica Acta* 44: 1543–1577.

Cairns-Smith AG (1971) *The Life Puzzle: On Crystals and Organisms and on the Possibility of a Crystal as an Ancestor*, p. 165. Toronto: University of Toronto Press.

Campbell DB, Campbell BA, Carter LM, Margot J-L, and Stacy NJS (2006) No evidence for thick deposits of ice at the lunar south pole. *Nature* 443: 835–837.

Campbell IB and Claridge GGC (1987) *Antarctica: Soils, Weathering Processes, and Environment*, p. 368. Amsterdam: Elsevier.

Campbell IH and Taylor SR (1983) No water, no granites – No oceans, no continents. *Geophysical Research Letters* 10: 1061–1064.

Carr MH (1981) *The Surface of Mars*, p. 232. New Haven, CT: Yale University Press.

Chekhov A (1899) *Uncle Vanya*, p. 59. London: Methuen (translation of 1986 by M. Frayn).

Cloud P (1976) Beginnings of biospheric evolution and their biogeochemical consequences. *Paleobiology* 2: 351–387.

Corliss JB, Baross JA, and Hoffman SE (1981) An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. In: *Oceanologica Acta, Proceedings of the 26th International Geology Congress*, Paris, pp. 59–69.

Crick FHC (1981) *Life Itself: Its Origin and Nature*, p. 192. New York: Simon and Schuster.

Darwin C (1959) Some unpublished letters. In: de Beer G (ed.) *Notes and Records of the Royal Society of London* 14: 12–66.

Dawkins R and Krebs JR (1979) Arms races between and among species. *Royal Society of London Proceedings* B205: 489–511.

Dietrich WE and Taylor Perron J (2006) The search for a topographic signature of life. *Nature* 439: 411–418.

Endress M, Zinner E, and Bischoff A (1996) Early aqueous activity on primitive meteorite parent bodies. *Nature* 379: 701–704.

Engel AE, Itson AP, Engel CG, Stickney DM, and Cray EJ (1974) Crustal evolution and global tectonics: A petrogenic view. *Geological Society of America Bulletin* 85: 843–858.

England GL, Rasmussen B, Krapez B, and Groves DI (2002) Palaeoenvironmental significance of rounded pyrite in siliciclastic sequences of the late Archaean Witwatersrand Basin: Oxygen-deficient atmosphere or hydrothermal alteration? *Sedimentology* 49: 1133–1156.

Ertem G and Ferris JP (1996) Synthesis of RNA oligomers on heterogeneous templates. *Nature* 379: 238–240.

Farquhar J, Bao H, and Thiemens M (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* 289: 756–758.

Farrow CEG and Mossman DJ (1988) Geology of Precambrian paleosols at the base of the Huronian Supergroup, Elliot Lake, Ontario, Canada. *Precambrian Research* 42: 107–139.

Feldman WC, Maurice S, Binder AB, Barraclough BL, Elphic RC, and Lawrence DJ (1998) Fluxes of fast and epithermal neutrons from Lunar Prospector: Evidence for water ice at the lunar poles. *Science* 281: 1496–1500.
Fogg MJ (1998) Terraforming Mars: A review of current research. *Advances in Space Research* 22: 415–420.
Folk RL and Lynch LE (1997) Nannobacteria are alive on Earth as well as Mars. *Society of Photo-Optical Instrumentation Engineers Proceedings* 3441: 112–122.
Formisano V, Atreya S, Encrenaz T, Ignatiev N, and Giuranna M (2004) Detection of methane in the atmosphere of Mars. *Science* 306: 1758–1761.
Forterre P (1995) Thermoreduction: An hypothesis for the origin of prokaryotes. *Comptes Rendus de l'Académie des Sciences Paris, Sciences et Vie* 318: 415–422.
Fowler AC and O'Brien SBG (2003) Lithospheric failure on Venus. *Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences* 459: 2663–2704.
Fox SL and Dose K (1972) *Molecular Evolution and the Origin of Life,* p. 359. San Francisco, CA: Freeman.
Furnes H, Banerjeee NR, Muehlenbachs K, Staudigel H, and de Wit M (2004) Early life recorded in Archean pillow lavas. *Science* 304: 578–581.
Garcia-Ruiz JM, Hyde ST, Carnerup AM, Christy AG, van Kranendonk MJ, and Welham NJ (2003) Self-assembled silica-carbonate structures and detection of ancient microfossils. *Science* 302: 1194–1197.
Gilbert GK (1877) *Report on the Geology of the Henry Mountains. US Geographical and Geological Survey of the Rocky Mountain Region (Powell Survey)*, p. 160. Washington, DC: Government Printing Office.
Goldich SS (1938) A study in rock weathering. *Journal of Geology* 46: 17–58.
Grotzinger JP (1988) Construction of early Proterozoic (1.9 Ga) barrier reef complex, Rocknest Platform, Northwest Territories. In: Geldsetzer HHJ, James NP, and Tebbutt GE (eds.) *Reefs, Canada and Adjacent Areas*, vol. 13, pp. 30–37. Calgary, AB, Canadian Society of Petroleum Geologists, Memoir.
Hallbauer DK, Jahns HM, and Beltmann HA (1977) Morphological and anatomical observations on some Precambrian plants from the Witwatersrand, South Africa. *Geologische Rundschau* 66: 477–491.
Halliday AN (2004) Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* 427: 505–509.
Hasiotis ST and Dubiel DL (1995) Termite (Insecta, Isoptera) nest ichnofossils from the Upper Triassic Chinle Formation, Petrified Forest National Monument. *Ichnos* 4: 111–130.
Heiken G, Morris RV, McKay DS, and Fruland RM (1976) Petrographic and ferromagnetic resonance studies of the Apollo 15 deep drill core. *Geochimica et Cosmochimica Acta* 4(supplement): 191–213.
Hole FD (1981) Effects of animals on soil. *Geoderma* 25: 75–112.
Horodyski RJ and Knauth LP (1994) Life on land in the Precambrian. *Science* 263: 494–498.
Itoh S and Yurimoto H (2003) Contemporaneous formation of chondrules and refractory inclusions in the early Solar System. *Nature* 423: 728–731.
Jellinek AM, Manga M, and Saar MO (2004) Did melting glaciers cause volcanic eruptions in eastern California? Probing the mechanics of dike formation. *Journal of Geophysical Research* 109(B9): 10.
Jull AJT, Courtney C, Jeffrey DA, and Beck JW (1998) Isotopic evidence for a terrestrial source of organic compounds found in the Martian meteorite Allan Hills 84001 and Elephant Moraine 79001. *Science* 279: 366–369.
Kerridge JF, Mackay AL, and Boynton WV (1979) Magnetite in CI carbonaceous chondrites: Origin by aqueous activity on a planetesimal surface. *Science* 205: 393–397.
Kirschvink JL and Kopp RE (2008) Palaeoproterozoic ice houses and the evolution of oxygen-mediating enzymes: The case for a late origin of photosystem II. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* B363: 2755–3765.
Knauth LP (1998) Salinity history of the Earth's early ocean. *Nature* 395: 554–555.
Kowalevsky V (1873) *Anchitherium aurelianense* Cuv. et sur l'histoire paleontologique des chevaux. *Mémoires de l'Academie Imperiale des Sciences de Saint-Petersbourg* 20: 1–73.
Kramer SN (1944) *Sumerian Mythology*, vol. 21, p. 25. Philadelphia, PA: American Philosophical Society, Memoir.
Lewis RS and Anders E (1975) Condensation time of solar nebula from extinct ^{125}I in primitive meteorites. *Proceedings of the National Academy of Sciences of the United States of America* 72: 268–271.
Lovelock JE (2000) *Gaia: The Practical Science of Planetary Medicine,* p. 192. London: Gaia Books.
MacFadden BJ (1992) *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae,* p. 369. Cambridge: Cambridge University Press.
MacFarlane AW, Danielson A, Holland HD, and Jacobsen SB (1994) REE chemistry and SM-Nd systematices of the Archaean weathering profiles in the Fortescue Group, Pilbara Block, Western Australia. *Geochimica et Cosmochimica Acta* 58: 1777–1794.
MacRae C (1999) *Life Etched in Stone,* p. 384. Johannesburg: Geological Society of South Africa.
Maher KA and Stevenson DJ (1988) Impact frustration of the origin of life. *Nature* 331: 612–614.
Marion GM, Kargel JS, Catling DC, and Jakubowski SD (2005) Effects of pressure on aqueous chemical equilibria at subzero temperatures with applications to Europa. *Geochimica et Cosmochimica Acta* 69: 259–274.
Maynard JB (1992) Chemistry of modern soils as a guide to interpreting Precambrian paleosols. *Journal of Geology* 100: 279–289.
McKay DS and Basu A (1983) The production curve for agglutinates in planetary regoliths. *Journal of Geophysical Research* 88(supplement): 193–199.
McKay DS, Gibson EK, Thomas-Kepeta KL, et al. (1996) Search for past life on Mars: Possible relic biogenic activity in Martian meteorite ALH84001. *Science* 273: 924–930.
McSween HY (1997) Evidence for life in a Martian meteorite? *GSA Today* 7(7): 1–7.
Meyer CE (1985) Ore metals through geologic history. *Science* 227: 1421–1428.
Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* 117: 528–529.
Mojzsis SJ, Harrison TM, and Pidgeon RT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409: 178–181.
Monger HC, Daugherty LA, Lindemann WC, and Liddell CM (1991) Microbial precipitation of pedogenic calcite. *Geology* 19: 997–1000.
Mossman DJ, Minter WEL, Dutkiewiczc A, et al. (2008) The indigenous origin of Witwatersrand "carbon". *Precambrian Research* 164: 173–186.
Navarro-González R, Rainey FA, Molina P, et al. (2003) Mars-like soils in the Atacama Desert, Chile, and dry limit of microbial life. *Science* 302: 1018–1021.
Nimmo F and Tanaka K (2005) Early crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161.
Nisbet EG and Sleep NH (2001) The habitat and nature of early life. *Nature* 409: 1083–1091.
Nishizawa M, Takahata N, Terada K, Komiya T, Ueno Y, and Sano Y (2005) Rare-earth element, lead, carbon, and nitrogen geochemistry of apatite-bearing metasediments from the approximately 3.8 Ga Isua supracrustal belt, West Greenland. *International Geology Review* 47: 952–970.
Palumbi SR (2001) Humans as the world's greatest evolutionary force. *Science* 293: 1786–1791.
Pelczar MJ, Chan ECS, Krieg NR, and Pelczar ME (1986) *Microbiology,* p. 918. New York: McGraw-Hill.
Penck W (1953) *Morphological Analysis of Land Forms: A Contribution to Physical Geology*, p. 429. London: Macmillan (translated by H. Czech and K.C. Boswell).
Pieri D (1999) Geomorphology of selected massifs on the Plains of Cydonia, Mars. *Journal of Scientific Exploration* 13: 401–412.
Pizzarello S, Huang Y, Becker L, et al. (2001) The organic content of the Tagish Lake meteorite. *Science* 293: 2236–2239.
Prashnowsky AA and Schidlowski M (1967) Investigation of Pre-Cambrian thucholite. *Nature* 216: 560–563.
Rech JA, Currie BS, Michalski G, and Cowan AM (2006) Neogene climate change and uplift in the Atacama Desert, Chile. *Geology* 34: 761–764.
Retallack GJ (1991) *Miocene Paleosols and Ape Habitats of Pakistan and Kenya,* p. 246. New York: Oxford University Press.
Retallack GJ (1997a) Early forest soils and their role in Devonian global change. *Science* 276: 583–585.
Retallack GJ (1997b) Palaeosols in the Upper Narrabeen Group of New South Wales as evidence of Early Triassic palaeoenvironments without exact modern analogues. *Australian Journal of Earth Sciences* 44: 185–201.
Retallack GJ (2001a) *Soils of the Past,* p. 600, 2nd edn. Oxford: Blackwell.
Retallack GJ (2001b) Cenozoic expansion of grasslands and global cooling. *Journal of Geology* 109: 407–426.
Retallack GJ (2001c) *Scoyenia* burrows from Ordovician paleosols of the Juniata Formation in Pennsylvania. *Palaeontology* 44: 209–235.
Retallack GJ (2002) Carbon dioxide and climate over the past 300 million years. *Philosophical Transactions of the Royal Society of London* A360: 659–674.
Retallack GJ (2004a) Late Oligocene bunch grassland and Early Miocene sod grassland paleosols from central Oregon, U.S.A. *Palaeogeography, Palaeoclimatology, Palaeoecology* 207: 203–237.
Retallack GJ (2004b) Soils and global change in the carbon cycle over geological time. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 5 (edited by T. Drever), pp. 581–605. Amsterdam: Elsevier Pergamon.
Retallack GJ (2005) Pedogenic carbonate proxies for amount and seasonality of precipitation in paleosols. *Geology* 33: 333–336.

Retallack GJ (2011) Marine influence in the Upper Ordovician Juniata Formation (Potters Mills, Pennsylvania): Implications for the history of life on land. *Palaios* 26: 675–677 (comment).
Retallack GJ (2012) Mallee model for Mesozoic and early Cenozoic mammalian communities. *Palaeogeography, Palaeoclimatology, Palaeoecology* 342–343: 111–129.
Retallack GJ and Huang C-M (2011) Ecology and evolution of Devonian trees in New York, USA. *Palaeogeography, Paleoclimatology, Paleoecology* 299: 110–128.
Retallack GJ and Mindszenty A (1994) Well preserved Late Precambrian paleosols from northwest Scotland. *Journal of Sedimentary Research* A64: 264–281.
Richardson SM (1978) Vein formation in the CI carbonaceous chondrites. *Meteoritics* 14: 141–159.
Rosing MT, Bird DK, Sleep NH, Glassley W, and Albarede F (2006) The rise of continents – An essay on the geologic consequences of photosynthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology* 232: 99–113.
Rosing MT and Frei R (2004) U-rich Archaean sea-floor sediments from Greenland: Indications of >3700 Ma oxygenic photosynthesis. *Earth and Planetary Science Letters* 217: 237–244.
Ross GM and Chiarenzelli P (1984) Paleoclimatic significance of widespread Proterozoic silcretes in the Bear and Churchill provinces of northwestern Canadian Shield. *Journal of Sedimentary Petrology* 55: 196–204.
Rowley DB (2002) Rate of plate creation and destruction: 180 Ma to present. *Geological Society of America Bulletin* 114: 927–933.
Rudnick RL (1995) Making continental crust. *Nature* 378: 571–577.
Rye R and Holland HD (1998) Paleosols and the evolution of atmospheric oxygen: A critical review. *American Journal of Science* 298: 621–672.
Rye R and Holland HD (2000) Life associated with a 2.76 Ga ephemeral pond? Evidence from Mount Roe #2 Paleosol. *Geology* 28: 483–486.
Sagan C and Pollack JB (1974) Differential transmission of sunlight on Mars: Biological implications. *Icarus* 21: 1217–1221.
Schieber J and Arnott HJ (2003) Nannobacteria as a by-product of enzyme-driven tissue decay. *Geology* 31: 717–720.
Schoonen M, Smirnov A, and Cohn C (2004) A perspective on the role of minerals in prebiotic synthesis. *Ambio* 33: 539–551.
Schopf JW (ed.) (1983) *Earth's Earliest Biosphere,* p. 543. Princeton: Princeton University Press.
Schopf JW (1999) Life on Mars: Tempest in a teapot? A first-hand account. *Proceedings of the American Philosophical Society* 143: 359–378.
Schopf JW and Klein C (eds.) (1991) *The Proterozoic Biosphere,* p. 1348. Cambridge: Cambridge University Press.
Schopf JW, Kudryatsev AB, Agresti DG, Wdowiak TJ, and Czaja AD (2002) Laser-Raman imagery of Earth's earliest fossils. *Nature* 416: 73–76.
Schumm SA (1977) *The Fluvial System,* p. 388. New York: Wiley.
Soil Survey Staff (2010) *Keys to Soil Taxonomy,* p. 338. Washington, DC: USDA-Natural Resources Conservation Service.
Squyres SW, Arvidson RE, Bell JF, et al. (2004) The Opportunity Rover's Athena science investigation at Meridiani Planum, Mars. *Science* 306: 1698–1703.
Stanley SM and Hardie LA (1999) Hypercalcification: Paleontology links plate tectonics and geochemistry to sedimentology. *GSA Today* 9(2): 1–7.
Stewart AD (2002) *The Late Proterozoic Torridonian Rocks of Scotland: Their Sedimentology, Geochemistry and Origin*, pp. 1–130. London: Geological Society, Memoir No. 24.
Strudley MW, Murray AB, and Haff PK (2006) Emergence of pediments, tors and piedmont junctions from a bedrock weathering-regolith thickness feedback. *Geology* 34: 805–808.
Taylor SR (1982) *Planetary Science: A Lunar Perspective,* p. 481. Houston, TX: Lunar and Planetary Institute.
Taylor G and Eggleton RA (2001) *Regolith Geology and Geomorphology,* p. 375. Chichester: Wiley.
Tice MM and Lowe DR (2004) Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* 431: 549–552.
Tomeoka K and Busek P (1988) Matrix mineralogy of the Orgueil CI carbonaceous chondrite. *Geochimica et Cosmochimica Acta* 52: 1622–1640.
Van Kranendonk MJ, Philippot P, Lepot K, Bodorkos S, and Pirajno F (2008) Geological setting of the Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. *Precambrian Research* 167: 93–124.
Van Valen L (1973) A new evolutionary law. *Evolutionary Theory* 1: 1–30.
Vermeij GJ (2004) *Nature: An Economic History,* p. 445. Princeton: Princeton University Press.
Verrecchia EP, Dumont J-L, and Verrecchia KE (1993) Role of calcium oxalate biomineralization by fungi in the formation of calcretes: A case study from Nazareth, Israel. *Journal of Sedimentary Petrology* 63: 1000–1006.
Veverka J, Thomas PC, Robinson M, et al. (2001) Imaging of small-scale features on 493 Eros from NEAR: Evidence for a complex regolith. *Science* 292: 484–488.
Wasson JT (1985) *Meteorites: Their Record of Early Solar-System History,* p. 267. New York: W.H. Freeman.
Watanabe Y, Martini JE, and Ohmoto H (2000) Geochemical evidence for terrestrial ecosystems 2.6 billion years ago. *Nature* 408: 574–578.
Watson AJ and Lovelock JE (1983) Biological homeostasis of the global environment: The parable of Daisyworld. *Tellus* 35B: 286–289.
Whitby J, Burgess R, Turner G, Gilmour J, and Bridges J (2000) Extinct ^{129}I in halite from a primitive meteorite: Evidence for evaporite formation in the early Solar System. *Science* 288: 1819–1821.
Whitfield J (2003) Alien-hunters get scope to search. *Nature* 422: 249.
Wood CA and Ashwal LD (1981) SNC meteorites: Igneous rocks from Mars? *Geochimica et Cosmochimica Acta* 12(supplement): 1359–1375.
Yen AS, Kim SS, Hecht MH, Frant MS, and Murray B (2000) Evidence that the reactivity of Martian soil is due to superoxide ions. *Science* 289: 1909–1912.
Zahnle K (2001) Decline and fall of the martian empire. *Nature* 412: 209–213.
Zhang F, Xu H, Konishi H, Shelobolina ES, and Roden EE (2012) Polysaccharide-catalyzed nucleation and growth of disordered dolomite: A potential precursor of sedimentary dolomite. *American Mineralogist* 97: 556–567.